

DEFINITION Sequence 38 from Patent WO0078968.
ACCESSION AX067463
VERSION AX067463.1 GI:12545083
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis.
Bacteria: Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 94750)
AUTHORS Lagace, R.E., Patterson, C. and Berg, K.L.
TITLE Nucleotide sequences of moraxella catarrhalis genome
JOURNAL Patent: WO 0078968-A 38 28-DEC-2000;
Incyte Genomics, Inc. (US)
FEATURES
Source 1. 94750
Location/Qualifiers
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 27128 a 18811 c 21444 g 27367 t
ORIGIN
alignment_scores:
Quality: 558.00 Length: 111
Ratio: 5.027 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
Alignment block:
US-10-048-197-2 x AX067463 ..
Align seg 1/1 to: AX067463 from: 1 to: 94750
1 MetLeuAnHsHsIleArgLeuThrIleSerAlaLeuThrAlaLe 17
|||||
75267 ATGTAATATCATATTCGCTCAGATATTCCTTACGACGCACT 75316
17 ULLeuValThrcIgcYsValSerThrgIyAsnValAlaMetLysGluGln 34
|||||
75317 TTTGCTAACAGCTGTGTTCTACTGTAATGTAATGCAAGCAAA 75366
34 sngIngnThrlIeGluGlnThrIleIleLysGlyThrAsnLysGln 50
|||||
75367 ACCAACAAACATTGACGACACCATCAATTAAGGCAAGCAACCA 75416
51 GluIleSerSerArgpHegIySerAlaAspSerIleSerPheMetIleVal 67
|||||
75417 GAGATTTCTACTAGATTTGCTTGGTATGATGATCTCTTTATGATAGT 75466
67 lValIleLysPheGlyHsThrAlaIleLeuAlaProAsnArgTrpGln 84
|||||
75467 GGTATATAAAATTTGACATACCCGCAATCTGCTCAAAACGATGCAAG 75516
84 lIleIleSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
|||||
75517 AAATTTATCCCTTATATTTCTTTCTTGGTGAACCATACAGACA 75566
101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
|||||
75567 AAGAACTTGCTATTTTATTTGACAGCAAGCA 75599
seq_name: gb_ba:AE005493
seq_documentation_block:
LOCUS AE005493 10701 bp DNA linear BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 112
of 290
ACCESSION AE005493 AE005174
VERSION AE005493.1 GI:12517067
KEYWORDS
SOURCE Escherichia coli O157:H7 EDL933.
ORGANISM Escherichia coli O157:H7 EDL933
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 10701)

AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
Grothbeck, E.J., Davis, N.W., Lin, A., Dimalanta, E., Potamousis, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Welch, R.A. and Blattner, F.R.
TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL Nature 409 (6819), 529-533 (2001)
MEDLINE 21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 10701)
AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
Grothbeck, E.J., Davis, N.W., Lin, A., Dimalanta, E., Potamousis, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Welch, R.A. and Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
Source
Location/Qualifiers
1. 10701
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:15864"
/note="enterohemorrhagic"
1. >10701
/note="O-island #108. Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
complement(224..829)
/gene="23935"
complement(224..829)
/gene="23935"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 82 to 187 of 201 are 30.84 pct identical
to residues 13 to 118 of 121 from Genpept 118 :
g113583|emb|CAA24404.1 (V01146) gene 3.8 [Bacteriophage
T7]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933y"
/protein_id="AAG57748.1"
/db_xref="GI:12517068"
/translation="MAKTRKPLRRKKAKAPARKPCNKRREDLAKRIFELSRDST
GVFRKVPYQGRALNSVAGAFDSNGYSIMIDGRYTHLVLFVITRNRPAGQIDH
VNGITDRPENLRCELPIDENSRNIRIKNSGCGGTWHRKOKNNVRLGFGKSK
HFGCPDDELAVLVAEARDKYGDSENSERYANLSEK"
1126..1881
/gene="trr8.3"
/note="23936"
1126..1881
/gene="trr8.3"
/function="putative enzyme; Transposases (Insertion
Sequence Associated)"
/note="Residues 1 to 247 of 251 are 100.00 pct identical
to residues 1 to 247 of 383 from Genpept 118
g11780028|gb|AAE9884.1|AF250878.45 (AF250878) IS30
transposase [Salmonella typhi]"
/codon_start=1
/transl_table=1
/product="IS30 transposase"
/protein_id="AAG57748.1"
/db_xref="GI:12517069"
/translation="MRTFTAEKASVVELKNGTSGSEIANILGSKPGTFTYLRDT
GKIRPHERKRAVAHLLTSEREIRIAGLSAKSISALATLNRSPSTISREVRGRGR
YKAVDANNRANRAKRPKPCLDONLPRLKYLEKLEKMSPOIGMLERTPDQR
TLRISPEITVIFRSRELAHNIHILRSHSLRGRNTRRGEGTINIVNGTIT
HERSRNIDRSLSGHWEDLVSGTKNSHISFP"
1941..2219
/gene="23937"
1941..2219
CDS
gene
CDS

THIS PAGE BLANK (USPTO)

Thu Sep 19 07:25:50 2002

OM of: US-10-048-197-2 to: GenEmbl:* out_format : pfs

Date: Sep 18, 2002 5:46 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+P2n.model -DEV=xlp
 -O=/ggn2.1/USFO_SPOOL/US10048197/runat_17092002.133849.22601/app-query.fasta.1.168
 -DB=GenEmbl -QFMT=fastlap -SUFFIX=std.rge -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
 -OGAPEXT=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bloms62
 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
 -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pts
 -NORR=ext -REAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USER=US10048197 -ECGN1_1_6377 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPRY -WAIT -THREADS=1

Search information block:

Query: US-10-048-197-2
 Query length: 111
 Database: GenEmbl:*
 Database sequences: 1797656
 Database length: 187333701
 Search time (sec): 1887.650000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pat:AX081233	+	558.00	1229.15	4.2e-60	336 AX081233 Sequence 1 from Patent
gb_pat:AX067463	+	558.00	1174.92	4.4e-57	94750 AX067463 Sequence 38 from Patent
gb_da:AE005493	+	121.50	228.56	0.0002	10701 AE005493 Escherichia coli O15
gb_da:AP002562	+	121.50	197.52	0.0122	270365 AP002562 Escherichia coli O1
gb_hlg:AC101573	+	81.50	121.33	214.55	73989 AC101573 Mus musculus clone R
gb_hlg:AC094388	+	81.50	115.13	475.06	140984 AC094388 Rattus norvegicus c
gb_hlg:AL627314	+	81.50	111.93	715.97	196632 AL627314 Mus musculus chromo
gb_hlg:AL627253	+	81.50	109.25	1.0e+03	259837 AL627253 Mus musculus chromo
gb_da:AF169455	+	80.50	173.18	0.2775	267 AF169455 Neisseria meningitidis
gb_pat:AX082148	+	80.50	173.18	0.2775	267 AX082148 Sequence 89 from Patent
gb_pat:AX024066	+	80.00	169.78	0.4294	339 AX024066 Sequence 9 from Patent
gb_pat:AX033695	+	80.00	169.78	0.4294	339 AX033695 Sequence 1 from Patent
gb_in:CEB04931	+	79.00	122.13	193.50	38240 CEB04931 Caenorhabditis elegans
gb_da:NM242491	+	79.00	100.88	3.0e+03	349061 NM242491 Homo sapiens BAC cl
gb_hlg:AC102510	+	78.50	115.88	431.56	65312 AC102510 Mus musculus clone R
gb_pat:BD003755	+	78.00	134.80	38.15	8136 BD003755 Polynucleotide of Str
gb_da:AE008532	+	78.00	132.44	51.59	10393 AE008532 Streptococcus pneum
gb_da:AE007477	+	78.00	130.41	66.96	12841 AE007477 Streptococcus pneum
gb_pat:AC012068	+	78.00	105.78	1.6e+03	166516 AC012068 Homo sapiens BAC cl
gb_hlg:AL135637	+	76.50	108.59	1.1e+03	87923 AL135637 Homo sapiens chr10M5
gb_hlg:AL138848	+	76.50	107.13	1.3e+03	102351 AL138848 Homo sapiens chromo
gb_pat:AL162430	+	75.50	100.14	3.2e+03	211791 AL162430 Human DNA sequence
gb_pat:AL137850	+	75.50	102.93	2.3e+03	125842 AL137850 Human DNA sequence
gb_hlg:AC011161	+	75.50	98.45	4.0e+03	200532 AC011161 Homo sapiens clone
gb_ov:AF173857	+	75.00	132.98	48.13	4920 AF173857 Xenopus laevis recep
gb_pat:AL512558	+	74.50	99.72	3.4e+03	139491 AL512558 Human DNA sequence
gb_hlg:AP001795	+	74.50	99.42	3.6e+03	143873 AP001795 Homo sapiens chromo
gb_p1:AP003215	+	74.50	98.76	3.9e+03	154137 AP003215 Oryza sativa genom
gb_hlg:AC099643	+	74.50	97.73	4.4e+03	171660 AC099643 Mus musculus clone
gb_pat:AC021986	+	74.50	97.70	4.4e+03	172182 AC021986 Homo sapiens clone
gb_pat:AC019184	+	74.50	96.53	5.2e+03	194372 AC019184 Homo sapiens BAC cl
gb_da:AF253043	+	74.50	95.99	5.5e+03	205646 AF253043 Lactobacillus crisp
gb_hlg:AC067965	+	74.00	143.01	13.30	1377 AC067965 Lactobacillus crisp
gb_da:AF421351	+	74.00	135.90	33.10	2884 AF421351 Acetobacter vinelan
gb_hlg:AC067965	+	74.00	106.08	1.5e+03	64165 AC067965 Arabidopsis thalian
gb_hlg:SPREU1902	+	74.00	97.79	4.4e+03	151947 SPREU1902 Streptococcus pneum
gb_hlg:AC009452	+	74.00	96.46	5.2e+03	174431 AC009452 Homo sapiens chromo
gb_hlg:AC099358	+	74.00	96.30	5.3e+03	174495 AC099358 Rattus norvegicus c
gb_pat:AL137020	+	74.00	96.04	5.5e+03	182231 AL137020 Human DNA sequence
gb_hlg:AC099386	+	74.00	94.36	6.8e+03	217121 AC099386 Rattus norvegicus c

seq_name: gb_pat:AX081233
 seq_documentation_block:
 LOCUS AX081233 336 bp DNA linear PAT 27-FEB-2001
 DEFINITION Sequence 1 from Patent WO0109337.
 ACCESSION AX081233
 VERSION AX081233.1 GI:13170101
 KEYWORDS
 SOURCE Moraxella catarrhalis.
 ORGANISM Moraxella catarrhalis.
 Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;

REFERENCE 1 (bases 1 to 336)

AUTHORS Thonard,J.S.

TITLE Moraxella catarrhalis antigens basp122 and basp124

JOURNAL Patent: WO 0109337-A 1 08-FEB-2001;

SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)

FEATURES Location/Qualifiers

source 1..336

/organism="Moraxella catarrhalis"

BASE COUNT 106 a 65 c 61 g 104 t

ORIGIN

alignment_scores:

Quality: 558.00 Length: 111
 Ratio: 5.027 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AX081233

Align seg 1/1 to: AX081233 from: 1 to: 336

1 MetLeuSnsHtSILeArgLeuThrILeSerAlaLeuThrAlaLe 17
 |||||||
 1 ATGTAAATCATCATATTCGGCTGACTATTCGGTTACGAGCGCACT 50
 |||||||
 17 UleuValThcIgcYsValSerThrgIAsnValAlaMetIcysGluGlnA 34
 |||||||
 51 TTGTGTACAGCTTGTCTTCTACTGCTAATGTTGCAATGAAGAGCAAA 100
 |||||||
 34 sngGngInThrILeGluInThrILeILeIcysGlyIcysThrAsnLysGln 50
 |||||||
 101 ACCAAACCAACCATGAGCAGACCATCATTAAGGCGAAGCAATAAGCAAA 150
 |||||||
 51 GluIleSerSerArgPheGlySerAlaAspSerILeSerPheMetIleLevA 67
 |||||||
 151 GAGATTTCCTACTAGATTGGTTCGTCGATAGCATCTCTTTATGATAGT 200
 |||||||
 67 lvalIleIcysPheGlyIcysThrAlaIleLeuAlaProAsnArgTrrGng 84
 |||||||
 201 GGTATTAATTAATTTGACATACCGCATCTCGCTCAACGATGCGCAAG 250
 |||||||
 84 lueIleSerLeuIleIleSerPheLeuTrpValIcysProTyrArgPro 100
 |||||||
 251 AAAATTTATCCCTATATATTTCTTTCTTTGGTGCAACCATACAGACCA 300
 |||||||
 101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
 |||||||
 301 AAGAACTTGTCATTTTATTTTACAGCAAGCA 333
 |||||||
 seq_name: gb_pat:AX067463
 seq_documentation_block:
 LOCUS AX067463 94750 bp DNA linear PAT 24-JAN-2001

```

DEFINITION Sequence 38 from Patent WO078968.
ACCESSION AX067463
VERSION AX067463.1 GI:12545083
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 94750)
AUTHORS Lagace,R.E., Patterson,C. and Berg,K.L.
TITLE Nucleotide sequences of Moraxella catarrhalis genome
JOURNAL Patent: WO 0078968-A 38 28-DEC-2000;
FEATURES
Source Location/Qualifiers
1..94750
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 27128 a 18811 c 21444 g 27367 t
ORIGIN

alignment_scores:
Quality: 558.00 Length: 111
Ratio: 5.027 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AX067463 ..

Align seg 1/1 to: AX067463 from: 1 to: 94750

1 MetLeuAnhShIsIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
75367 ATGTTAAATCATCATATTCGCCGACTATTTCGCTTATCTACGCACT 75316
17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGln 34
75317 TTTCGTAACAGGTTGGTGTCTTCTACGTAATGTTGCAATGAAGACAA 75366
34 sngInglThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
75367 ACCAACAACCATGAGCAGACCATCATTTAAGGCAAGACCAATGAAGCA 75416
51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVal 67
75417 GAGATTTCAGTAGATTTGGTCTCTGATGATGATCTCTTTATGATAGT 75466
67 ValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTyrGln 84
75467 GGTAAATAAATTGGACATACCGCATRACCTGCTCCAAACCGATGGCAAG 75516
84 LuIleLeuSerLeuIleIleSerPheLeuTyrValLysProTyrArgPro 100
75517 AATATTTCCTTATTAATTTCTTTCTTGGGGAACCATGACAGCA 75566
101 LysAnLeuSerPheTyrLeuThrAlaLysAla 111
75567 AAGACTTGTCATTTTATTGACAGCAAGCA 75599

seq_name: gb_ba:AE005493
seq_documentation_block:
LOCUS AE005493 10701 bp DNA linear BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 112
of 290.
ACCESSION AE005493 AE005114
VERSION AE005493.1 GI:12517067
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 10701)
Escherichia coli O157:H7 EDL933.
Escherichia coli O157:H7 EDL933.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

AUTHORS
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postfal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grodeck,E.J., Davis,N.W., Lim,A., Dimantanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
JOURNAL
MEDLINE
PUBMED
21074935
11206551
2 (bases 1 to 10701)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postfal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grodeck,E.J., Davis,N.W., Lim,A., Dimantanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
JOURNAL
FEATURES
Source Location/Qualifiers
1..10701
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
<1..>10701
/note="O-island #108; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
complement(224..829)
/gene="23935"
complement(224..829)
/gene="23935"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 82 to 187 of 201 are 30.84 pct identical
to residues 13 to 118 of 121 from Genpept 118 :
T71155831[emb]CA24404.1] (V01146) gene 3.8 [Bacteriophage
T7]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933y"
/protein_id="AAG57748.1"
/db_xref="GI:12517068"
/transl_table="MAKERKPLRRKKAMAKPARRCNRRKREDLVKRIFFLSFDKST
GVFRMKVPTQGRIALNSVAGAFDSNGYSIMIDRRKTYHYLYVYTHHNPFGQIDH
VNGIRTDNRPENLRCLPIENSRNIRIKNSGRCVTHAKRKQKMNVRGHRGSK
HRCFDDLELAVLVEARDKYYGDFSGNSRSTYANLSKEM"
1126..1881
/gene="trab.3"
/note="23936"
1126..1881
/gene="trab.3"
1126..1881
/gene="trab.3"
/function="putative enzyme; Transposases (Insertion
Sequence Associated)"
/note="Residues 1 to 247 of 251 are 100.00 pct identical
to residues 1 to 247 of 383 from Genpept 118 :
g1178002881[gb]AA69884.1[AF250878_45 (AF250878) IS30
transposase (Salmonella typhi)]"
/codon_start=1
/transl_table=11
/product="IS30 transposase"
/protein_id="AAG57749.1"
/db_xref="GI:12517069"
/transl_table="MRRTPAEKASVPELKNGTGFSELIANILSGKPTFTYMLRDT
GGIKPERRKAVAHILTSRETRRGLSKMSIRAIATALNRSPSTISREVOARGRR
TYKAVDNNRANRMAKRPKCLDQNLPLRKLVLEIKLWKMSPEQISGMLRTPKDRK
TLRISPTETITGLVFRSREALHLNIOHLRSHSLRHGRHRTTRGEGGTINIVNGTIT
HERSNIDNRSLGVEGDLVSGTNSHIDSF"
1941..2219
/gene="23937"
1941..2219
CDS
gene
CDS

```

```

/gene="23937"
/function="orf: Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/protein_id="AAG57750.1"
/db_xref="GI:12517070"
/translation="MKTRIYAGKTTKQDVLAFCGPPDSRLIDGEGMSYMYNSQSK
ATSFIPVGLAGADSOFTSLTVSEFGKEVSTYIFNAGTSNVTGIF"
2390..2509
/gene="23938"
2390..2509
/gene="23938"
/function="orf: Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/protein_id="AAG57751.1"
/db_xref="GI:12517071"
/translation="MALALALALALALALALALALALALALALALAKEN"
complement(2609..3517)
/gene="23939"
complement(2609..3517)
/gene="23939"
/function="orf: Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/protein_id="AAG57752.1"
/db_xref="GI:12517072"
/translation="MXIAKDSSTDEIHGNDLRGMDDEYIKTSFECPEYBCKIKAP
CSFTMKHVNOSYFRYGDHKDGCGLHDPYKNNHTSNDERKHNSEPAVLSLKIDV
PRGCVKARNKNNKNEHDEKKEHNEPVSSSIKPVVDYIINNSNNEOLSTIPYGTSS
YKDFPOLIFVKNNIRYKYPAYVGAOSNIRLHDSOHCCTPLARDKKQKPTLETI
DVSDMKSQKDYEPKYEKQKREADRTYGLKDKRNKKYLTFFFCAPDEKDFLEK
TNHFKLYVAFLGKFESSYNDSSNYH"
complement(3914..4384)
/gene="23940"
complement(3914..4384)
/gene="23940"
/function="orf: Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/protein_id="AAG57753.1"
/db_xref="GI:12517073"
/translation="MLFTFLSKQNFETIARHIMDKTNINLMERYLLLLDRFVDKLA
SGFSEORLIEQSYLCAGFYIKQSGIEKMTFSREVVLTFLLSYSHINKLDDLI
NKERMKHYCCSLINFIIVSNGSRTEKYVNERKYEASTLKRSLSKDKRRYRL"
complement(4425..4646)
/gene="23941"
complement(4425..4646)
/gene="23941"
/function="orf: Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/protein_id="AAG57754.1"
/db_xref="GI:12517074"
/translation="MKRTIKSGVIHTVKNPANKLVASQRIKRLAEANEALKALMERYL
EOPFVLYNAHLKOITIELNNPLPAKYL"
complement(5046..6648)
/gene="23942"
complement(5046..6648)
/gene="23942"
/function="orf: Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"

```

```

/codon_start=1
/transl_table=1
/protein_id="AAG57755.1"
/db_xref="GI:12517075"
/translation="MNNIFIFEPSKNNPNDLVNIKFEFCSTISNNLTTSWESNKK
KGLRYFTFKNSKNNLSKEDPSFINAKYVMHLVHSEFNKSKTKHSLMLKIVEFY
LTKINMEANVYCNNSIDECIRIASEKYSAAHFAIKKELEKSSFTLNDNRMNSEFY
LFWNPRIYRTQSWTGYDSSLBGSRLPDIKSYIAIAIEFSKDEOLSRDFTTSY
LALMLCAPSRISSEIALADPACETIECDCKGIQRIQGLRFSKKGEGNTKAPITLAIY
AKKAISRLKEKESQARLAEIIONKYSNSTGTLENTIPDLFWYDRKRIKYSNALC
LITTEGOLNOKKKEMSDKLFRPTTFEFTDIIDSDYIKGFVFWFKRHGIYINDGSYLL
XTHOLRHLNLTFAIXNGMDERSIARMSGRKLISQNVSDRSHLQMSKAIREKKLSVC
VNEHRIDIPVVDLNEFGLT"
complement(6455..7780)
/gene="23943"
complement(6455..7780)
/gene="23943"
/function="putative enzyme: Integration, recombination
(Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/protein_id="AAG57756.1"
/db_xref="GI:12517076"
/translation="MEPSMENKCIESPQIFPAKNRYSFKISDKKMOLEKNCYPPK
VYDRAPTKMSLYKTLAYASEVSSFTIOSINNLFYEMFCAMTIDTIDKAIYOLN
YLGSERNKMLIKAFIIRKNNLPGEVAIRMLEIKIIPQOTGAVARRDNRKG
PLTEAEFNININAVGFYHEKKIOCFILYLLAIIRLQILSLAKDLINKERG
CFYLVKPKYKORCKFERFENFMVMEIFLTDLSMLINQAEVEDEKSYGIDNYPELP
IFMNDKITEETRIEDFLYDLTDFEHLKNSVMSLKLKPKSPDVRSEPTNSYELN
ARRPRYTGSRLANGASIEVIYAKALDHKSNSSTIYKKNPDVYDIDKRLSNFNP
LSNIIIGIEIEENKNFFIKFVSDAFPLEEDTKEDKLCITCKKFNPRAL"
complement(7935..8978)
/gene="23945"
complement(7935..8978)
/gene="23945"
/function="putative enzyme: Integration, recombination
(Phage or Prophage Related)"
/note="Residues 172 to 336 of 347 are 28.57 pct identical
to residues 125 to 264 of 311 from GenPept 118 :
g126219881gb|AAB85391.1 (AE000865) integrase-recombinase
protein [Methanobacterium thermoautotrophicum]"
/codon_start=1

```

alignment_scores:

Quality	Ratio	Length	Gaps
121.50	1.960	93	2
Percent Similarity: 66.667	Percent Identity: 34.409		

alignment_block:

us-10-048-197-2 x AE005493 ..

Align seg 1/1 to: AE005493 from: 1 to: 10701

```

17 LeuLeuValThrGlyCysValSerThrGlyAsnValAlaMetGluGlu 33
   || ..... ||||| ||||| ||||| .....
1878 CTTACCTTGGCAGGTTGCTTTCAGTAGGCAACGACGATAAAATGA 1927
   ||||| ..... ||||| ||||| ||||| .....
33 nasngInglInThrIleGluInThrIleIleIleGlyIleIleGly 50
   ||||| ..... ||||| ||||| ||||| .....
1928 GAGCGAGGAAGTGTGAAACCAAAATTTTAAAGCCAAACCAACAAAC 1977
   ||||| ..... ||||| ||||| ||||| .....
50 InGluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIle 66
   ||||| ..... ||||| ||||| ||||| .....
1978 AGACGCTGTACGATGCTTCGGAACCTGACACCGCTTCT...TTGATC 2024
   ||||| ..... ||||| ||||| ||||| .....
67 ValValIleIleGlyPheGlyHisThrAlaIleLeuAlaProAsnArgTyr 83
   ||||| ..... ||||| ||||| ||||| .....
2025 GATGTGTGAACAAATGTCATATACATGTAACGACGACCAACG 2074
   ||||| ..... ||||| ||||| ||||| .....
83 InGluIleLeuSerLeuIleIleSerPheLeuTyrValIleProTyrArg 99
   ||||| ..... ||||| ||||| ||||| .....
2075 AACCTTTCATCCCGCTTGTGGGACTGCTTGCAAGTGGCGGACACTCAC 2124

```

```

100 ProlYasnLeuSerPheTyrLeuThr 108
|||||
2125 AACGAAATCTGACAGTTCTTTCA 2151
seq_name: gb_ba:AP002562
seq_documentation_block:
LOCUS AP002562 270365 bp DNA linear BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 13/20.
ACCESSION AP002562 BA000007
VERSION AP002562.1 GI:13362858
KEYWORDS
SOURCE
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DML.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (sites)
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Yamamoto,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Sasakawa,C. and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL MEDLINE 20198780
REFERENCE
2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
JOURNAL MEDLINE 20557356
REFERENCE
3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Matanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
JOURNAL MEDLINE 20564182
REFERENCE
4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tohe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Shinagawa,H., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
JOURNAL MEDLINE 21156231
REFERENCE
5 (bases 1 to 270365)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:kengen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp,
Fax:81-6-6879-2047)
Tel:81-6-6879-8365,
genome project.
COMMENT
FEATURES
Location/Qualifiers
1..270365
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
gene
CDS
complement(435..1220)
/gene="ECS3388"
complement(435..1220)
/gene="ECS3388"
/note="similar to SEB_ECOLI g11178871 percent identity
98 in 261 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="enhanced serine sensitivity"
/protein_id="BAB36811.1"
/db_xref="GI:13362859"
/translation="MYLSEKNELEDLEKATPEAPPAFFTELESTVWPGTA
OGEAVVDSALDIQHWKEDGTSVIPFFSLAQAVEDQAPVWVPTLFMTIG
LTIPLNKKLPFGKPEMPRETSLSIGENPLSSOELIGESGLITSEVAEPQMDIS
PIDICVYKKGKGISHTETHTAPFERRWGSFLDPQNRHIT"
complement(1354..2724)
/gene="ECS3389"
complement(1354..2724)
/gene="ECS3389"
/note="similar to PEPB_ECOLI g1178872 percent identity
99 in 456 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative peptidase"
/protein_id="BAB36812.1"
/db_xref="GI:13362860"
/translation="MPPSCNNAHAEKSMGCOALALORRTKMTKMTKSTOPAD
ARWGEKATYSINNDGTLILNAGADDLGLQARKIDGIGIHVQLSGMDADRCRA
DLISNVASDPRVYRTIKGEDLREQGVGLHTVGRGSEPVLAIDVPTDCKAPY
ACLYKGTGDSGGSTIKQYAFDMSKSDMGAAVYGLAFLATIRGNKVKYKLFLLC
ADNLTSGWAFRLDIIYRNKGKVEVYVNTDAGRLVADGLIDASQKPEMTIDAATL
TGAKKTALGNDYHLEFSDALAGRLLASQSENPFRPLDPAEHRNQLPSNFAELN
ANLITAA"
complement(2815..3015)
/gene="ECS3390"
complement(2815..3015)
/gene="ECS3390"
/note="identical to YFHU_ECOLI g11788873 (Conserved in
E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB36813.1"
/db_xref="GI:13362861"
/translation="WGLKWTDSREIGALYDAYPDLPKTVVFTDWHQWICDLEPFD
DPOASNEKILRALILVWDEAE"
complement(3027..3362)
/gene="ECS3391"
complement(3027..3362)
/gene="ECS3391"
/note="similar to FER_ECOLI g11199991sp125528 percent
identity 100 in 111 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="2Fe-2S ferredoxin"
/protein_id="BAB36814.1"
/db_xref="GI:13362862"
/translation="MKPIVILPHDLCPDGAVLEANGSETTLDAALNGLIEIHACEK
SCATTCCHCIVREGFDSLPESSSEQDDMLDKAWGLEPESRILSCQARTVEDLVLEIPR
YVTHNAREH"
complement(3364..5214)
/gene="ECS3392"
complement(3364..5214)
/gene="ECS3392"
/note="similar to HSCA_ECOLI g11788875 percent identity
100 in 616 aa (Conserved in E.coli K-12)"

```

[illegible]

seq_documentation_block:

LOCUS AC101573 73989 bp DNA linear HTG 23-NOV-2001
DEFINITION Mus musculus clone RP23-197A22, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC101573
VERSION AC101573.1 GI:17060348
KEYWORDS HTG; HTGS; PHASEO.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE 1 (bases 1 to 73989)
JOURNAL Mus musculus, clone RP23-197A22
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 73989)

Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barina,N., Bastien,P.V., Boguslavsky,L., Bouknighter,B.,
Brown,A., Camarata,J., Campotondo,A., Chang,Y.J., Chazaro,B.,
Chapel,V., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., Dextrallano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,G., Gardyna,S.,
Grinde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hefford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamasz,R., Landers,T., Lehoczy,J., Levine,N., Matthews,C.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Meltzer,J.,
McCarthy,M., McKean,P., McKernan,K., McNeeters,R., Meldrum,J.,
Menues,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nobdu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhan,P., Plette,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rice,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Strassman,K.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K.,
Travers,M., Travis,A., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zemke,
L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIR
Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center Project name: L17015
Center clone name: 197_A_22

* NOTE: This record contains 92 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```
1 719: contig of 719 bp in length
* 720 819: gap of 100 bp
* 820 1536: contig of 717 bp in length
* 1537 1636: gap of 100 bp
* 1637 2339: contig of 703 bp in length
* 2340 2439: gap of 100 bp
* 2440 3132: contig of 693 bp in length
* 3133 3232: gap of 100 bp
* 3233 3934: contig of 702 bp in length
* 3935 4034: gap of 100 bp
* 4035 4745: contig of 711 bp in length
* 4746 4845: gap of 100 bp
* 4846 5546: contig of 701 bp in length
* 5547 5646: gap of 100 bp
* 5647 6348: contig of 702 bp in length
* 6349 6448: gap of 100 bp
* 6449 7166: contig of 718 bp in length
* 7167 7266: gap of 100 bp
* 7267 7983: contig of 717 bp in length
* 7984 8083: gap of 100 bp
* 8084 8786: contig of 703 bp in length
* 8787 8886: gap of 100 bp
* 8887 9605: contig of 719 bp in length
* 9606 9705: gap of 100 bp
* 9706 10386: contig of 691 bp in length
* 10397 10496: gap of 100 bp
* 10497 11200: contig of 704 bp in length
* 11201 11300: gap of 100 bp
* 11301 12012: contig of 712 bp in length
* 12013 12112: gap of 100 bp
* 12113 12815: contig of 703 bp in length
* 12816 12915: gap of 100 bp
* 12916 13614: contig of 699 bp in length
* 13615 13714: gap of 100 bp
* 13715 14428: contig of 714 bp in length
* 14429 14528: gap of 100 bp
* 14529 15243: contig of 715 bp in length
* 15244 15343: gap of 100 bp
* 15344 16072: contig of 729 bp in length
* 16073 16172: gap of 100 bp
* 16173 16890: contig of 718 bp in length
* 16891 16990: gap of 100 bp
* 16991 17695: contig of 705 bp in length
* 17696 17795: gap of 100 bp
* 17796 18482: contig of 687 bp in length
* 18483 18582: gap of 100 bp
* 18583 19294: contig of 712 bp in length
* 19295 19394: gap of 100 bp
* 19395 20106: contig of 712 bp in length
* 20107 20206: gap of 100 bp
* 20207 20911: contig of 705 bp in length
* 20912 21011: gap of 100 bp
* 21012 21711: contig of 700 bp in length
* 21712 21811: gap of 100 bp
* 21812 22516: contig of 705 bp in length
* 22517 22616: gap of 100 bp
* 22617 23310: contig of 694 bp in length
*
* 23311 23410: gap of 100 bp
* 23411 24116: contig of 706 bp in length
* 24117 24216: gap of 100 bp
* 24217 24921: contig of 705 bp in length
* 24922 25021: gap of 100 bp
* 25022 25734: contig of 713 bp in length
* 25735 25834: gap of 100 bp
* 25835 26548: contig of 714 bp in length
* 26549 26648: gap of 100 bp
* 26649 27355: contig of 707 bp in length
* 27356 27455: gap of 100 bp
* 27456 28147: contig of 692 bp in length
* 28148 28247: gap of 100 bp
* 28248 28959: contig of 712 bp in length
* 28960 29059: gap of 100 bp
* 29060 29755: contig of 696 bp in length
* 29756 29855: gap of 100 bp
* 29856 30561: contig of 706 bp in length
* 30562 30661: gap of 100 bp
* 30662 31372: contig of 711 bp in length
* 31373 31472: gap of 100 bp
* 31473 32164: contig of 692 bp in length
* 32165 32264: gap of 100 bp
* 32265 32940: contig of 676 bp in length
* 32941 33040: gap of 100 bp
* 33041 33757: contig of 717 bp in length
* 33758 33857: gap of 100 bp
* 33858 34568: contig of 711 bp in length
* 34569 34668: gap of 100 bp
* 34669 35379: contig of 711 bp in length
* 35380 35479: gap of 100 bp
* 35480 36198: contig of 719 bp in length
* 36199 36298: gap of 100 bp
* 36299 37004: contig of 706 bp in length
* 37005 37104: gap of 100 bp
* 37105 37810: contig of 706 bp in length
* 37811 37910: gap of 100 bp
* 37911 38582: contig of 672 bp in length
* 38583 38682: gap of 100 bp
* 38683 39384: contig of 702 bp in length
* 39385 39484: gap of 100 bp
* 39485 40168: contig of 684 bp in length
* 40169 40268: gap of 100 bp
* 40269 40963: contig of 695 bp in length
* 40964 41063: gap of 100 bp
* 41064 41765: contig of 702 bp in length
* 41766 41865: gap of 100 bp
* 41866 42570: contig of 705 bp in length
* 42571 42670: gap of 100 bp
* 42671 43368: contig of 698 bp in length
* 43369 43468: gap of 100 bp
* 43469 44188: contig of 720 bp in length
* 44189 44288: gap of 100 bp
* 44289 44984: contig of 696 bp in length
* 44985 45084: gap of 100 bp
* 45085 45801: contig of 717 bp in length
* 45802 45901: gap of 100 bp
* 45902 46588: contig of 687 bp in length
* 46589 46688: gap of 100 bp
* 46689 47360: contig of 672 bp in length
* 47361 47460: gap of 100 bp
* 47461 48180: contig of 720 bp in length
* 48181 48280: gap of 100 bp
* 48281 48967: contig of 687 bp in length
* 48968 49067: gap of 100 bp
* 49068 49770: contig of 703 bp in length
* 49771 49870: gap of 100 bp
* 49871 50560: contig of 710 bp in length
* 50561 50680: gap of 100 bp
* 50681 51408: contig of 728 bp in length
* 51409 51508: gap of 100 bp
* 51509 52241: contig of 733 bp in length
* 52242 52341: gap of 100 bp
```

```

*      52342      53050: contig of 709 bp in length
*      53051 53150: gap of 100 bp
*      53151 53665: contig of 715 bp in length
*      53666 53965: gap of 100 bp
*      53966 54654: contig of 689 bp in length
*      54655 54754: gap of 100 bp
*      54755 55466: contig of 712 bp in length
*      55467 55566: gap of 100 bp

```

```

alignment_scores:
  Quality: 81.50      Length: 77
  Ratio: 1.598      Gaps: 1
  Percent Similarity: 66.234      Percent Identity: 27.273

```

```

alignment_block:
  US-10-048-197-2 x AC101573 ..

```

```

Align seg 1/1 to: AC101573 from: 1 to: 73989

```

```

31  LysGluGlnAnGlnGlnThrIleGluGlnThrIleIleLysGlyLysTh 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61496  AAAAAAAAAAACCAACCTCCATGCTTCCTTCCTTCCTTCCTTCCTTC 61545
47  rAnLysGlnLysSerArpPheGlySerAlaAspSerIleSerP 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61546  TAATTCGCCCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGT 61595
64  hemeIleValIleLysPheGlyHisThrAlaIleLeuAlaProAsn 80
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61596  TCGTTGTGTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 61645
81  ArGTPrGlnGlnIleLeuSerLeuIleIleSerPheLeuTrpAlaLysPr 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61646  .....TCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 61686
97  oTyrArGPrGlnAsnLeuSerPheTyrLeu 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61687  CTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 61717

```

```

seq_name: gb_hhg:AC094388

```

```

seq_documentation_block:
LOCUS      AC094388              140984 bp      DNA      linear      HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-3N14, *** SEQUENCING IN PROGRESS ***,
ACCESSION AC094388
VERSION    AC094388.2 GI:17941111
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

```

```

REFERENCE
AUTHORS    1 (bases 1 to 140984)
            Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C.,
            Albrooks S.L., Amaratunga H.C., Are J.R., Banks T., Barbara J.,
            Benton J., Bimberg K., Blankenburg K., Bonnin D., Bouck J.,
            Bowle S., Birkett M., Brown E., Brown M., Bryant N.P., Buhay C.,
            Burch P., Burkett C., Burrell K.L., Byrd N.C., Carron T.F.,
            Carter M., Cavazos S.R., Chacko J., Chavez D., Chen R.,
            Chen Z., Chowdhury I., Christopoulos C., Cleveland C.D., Cox C.,
            Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C.,
            Davy Carroll L., Dederich D.A., Delaney K.R., Delgado O.,
            Demn A.L., Ding Y., Dinh H.H., Douthwaite K.J., Draper H.,
            Dugan-Hochia S., Durbin K.J., Eamhart C., Edgar D., Edwards C.C.,
            Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J.,
            Foster P., Frantz P., Gablaj A., Gao J., Garcia A., Garner T.,
            Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S.,
            Hamilton K., Harris C., Harris K., Hart M., Havlak P., Hawes A.,
            Hernandez J., Hernandez O., Hodgson A., Hogue M., Hollway C.,
            Hollins B., Homs F., Howard S., Huber J., Hulyk S., Hume J.,
            Jackson L.E., Jacobson B., Jia Y., Johnson R., Jolivet S.,
            Joudah S., Karlsson E., Kelly S., Khan U., King L., Korvah J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

```

COMMENT

```

```

Kovar C., Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C.,
Lewis L., Li J., Li Z., Lichtenage O., Lieu C., Liu J., Liu W.,
LouiSeged H., Lozano R.J., Lu X., Lucier A., Lucier R., Lunn R.,
Ma J., Maheshwari M., Mapua P., Martin R., Martindale A.,
Martinez E., Massey E., Mawhney E., McLeod M.P., Meador M.,
Mel G., Metzker M., Miner G., Miner Z., Mitchell T., Mohabbat K.,
Morgan M., Morris S., Moser M., Neal D., Newton J., Newton N.,
Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nwokemwo S.,
Ogun M., Okwundu G., Oragunye N., Oviedo R., Pace A., Payton B.,
Peery J., Perez L., Peters L., Pickens R., Primus E., Pu L.L.,
Quiles M., Ren Y., Rives M., Rojas A., Rojupokan I., Rolfe M.,
Ruiz S., Savery G., Scherer S., Scott G., Shen H., Shoohtari N.,
Sisson I., Sodergren E., Sonalke T., Sparks A., Stanley H.,
Stone H., Sutton A., Swalek A., Taber P., Tamerisa A., Tamerisa K.,
Tang H., Tansey J., Taylor C., Taylor T., Telifod B., Thomas N.,
Thomas S., Usmani K., Vasquez L., Vera V., Villalón D., Vinson R.,
Wall R., Wang S., Ward-Moore S., Warren R., Washington C.,
Watlington S., Williams G., Williamson A., Wleczky R., Wooden S.,
Worley K., Wu C., Wu Y., Wu Y.F., Zhou D., Zorrilla S., Nelson D.,
Weinstock G. and GIBBS R.
Direct Submission
Unpublished
2 (bases 1 to 140984)
Morley K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624222.
----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GAMS
Center clone name: CH230-3N14
----- Summary Statistics -----
findphraplist
Assembly program: Phrap; version 0.990329First call to
Consensus quality: 112041 bases at least Q40
Consensus quality: 120432 bases at least Q30
Consensus quality: 127286 bases at least Q20
Estimated insert size: 121178; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-rip estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 3813: contig of 3813 bp in length
* 3814
* 3913: gap of unknown length
* 3914
* 8564: contig of 4651 bp in length
* 8565
* 8664: gap of unknown length
* 8665
* 12409: contig of 3745 bp in length
* 12410
* 12509: gap of unknown length
* 12510
* 16844: contig of 4335 bp in length
* 16845
* 16944: gap of unknown length
* 16945
* 21435: contig of 4491 bp in length
* 21436
* 21535: gap of unknown length
* 21536
* 23974: contig of 2439 bp in length
* 23975
* 24074: gap of unknown length
* 24075
* 28394: contig of 4330 bp in length
* 28395
* 28495: gap of unknown length
* 28496
* 33266: contig of 4772 bp in length
* 33267
* 33366: gap of unknown length

```



```

* 3367 36620: contig of 3254 bp in length
* 36621 36720: gap of unknown length
* 36721 36721: contig of 2895 bp in length
* 36716 39715: gap of unknown length
* 39716 42700: contig of 2985 bp in length
* 42701 42800: gap of unknown length
* 42801 46577: contig of 3777 bp in length
* 46578 46578: gap of unknown length
* 46578 48925: contig of 2248 bp in length
* 48926 49025: gap of unknown length
* 49026 51537: contig of 2511 bp in length
* 51537 51636: gap of unknown length
* 51637 54619: contig of 2983 bp in length
* 54620 54720: gap of unknown length
* 54720 57076: contig of 2356 bp in length
* 57076 57175: gap of unknown length
* 57175 59447: contig of 2272 bp in length
* 59448 59547: gap of unknown length
* 59548 63146: contig of 3598 bp in length
* 63146 63245: gap of unknown length
* 63246 66267: contig of 3022 bp in length
* 66268 66367: gap of unknown length
* 66368 69215: contig of 2848 bp in length
* 69216 69315: gap of unknown length
* 69316 72092: contig of 2777 bp in length
* 72093 72192: gap of unknown length
* 72193 73295: contig of 1103 bp in length
* 73296 73395: gap of unknown length
* 73396 75943: contig of 2548 bp in length
* 75944 76043: gap of unknown length
* 76044 78937: contig of 2893 bp in length
* 78937 79036: gap of unknown length
* 79037 80901: contig of 1865 bp in length
* 80902 81002: gap of unknown length
* 81003 83288: contig of 2187 bp in length
* 83289 85977: gap of unknown length
* 85978 86076: contig of 2688 bp in length
* 86077 87701: gap of unknown length
* 87702 87870: contig of 1694 bp in length
* 87871 88938: gap of unknown length
* 88939 91036: contig of 1068 bp in length
* 91037 91136: gap of unknown length
* 91137 93364: contig of 1998 bp in length
* 93365 93464: gap of unknown length
* 93465 94848: gap of unknown length
* 94849 94947: contig of 1383 bp in length
* 94948 96551: gap of unknown length
* 96552 96651: contig of 1604 bp in length
* 96652 98298: gap of unknown length
* 98299 98398: contig of 1647 bp in length
* 98399 100654: gap of unknown length
* 100655 100754: contig of 2256 bp in length
* 100755 101948: gap of unknown length
* 101949 102048: contig of 1194 bp in length
* 102049 103515: gap of unknown length
* 103516 103615: contig of 1467 bp in length
* 103616 105211: gap of unknown length
* 105212 105311: contig of 1596 bp in length
* 105312 106335: gap of unknown length
* 106336 106435: contig of 1124 bp in length
* 106436 106535: gap of unknown length
* 106536 107844: contig of 1209 bp in length
* 107845 107845: gap of unknown length
* 107845 109563: contig of 1719 bp in length
* 109564 109663: gap of unknown length
* 109664 112050: contig of 2387 bp in length
* 112051 112150: gap of unknown length
* 112151 113412: contig of 1262 bp in length
* 113413 113512: gap of unknown length
* 113513 114769: contig of 1257 bp in length
* 114770 114869: gap of unknown length
* 114870 116333: contig of 1464 bp in length

```

```

* 116334 116433: gap of unknown length
* 116434 117929: contig of 1496 bp in length
* 117930 118029: gap of unknown length
* 118030 119642: contig of 1613 bp in length
* 119643 119742: gap of unknown length
* 119743 121418: contig of 1676 bp in length
* 121419 121518: gap of unknown length
* 121519 123314: contig of 1795 bp in length
* 123314 123413: gap of unknown length
* 123414 124693: contig of 1280 bp in length
* 124694 124793: gap of unknown length
* 124794 126213: contig of 1420 bp in length
* 126214 126313: gap of unknown length
* 126314 128131: contig of 1818 bp in length
* 128132 128231: gap of unknown length
* 128232 128941: contig of 1710 bp in length
* 128942 130041: gap of unknown length
* 130042 131544: contig of 1503 bp in length
* 131545 131644: gap of unknown length
* 131645 133140: contig of 1496 bp in length
* 133141 133240: gap of unknown length
* 133241 134535: contig of 1296 bp in length
* 134536 134537: gap of unknown length

```

alignment_scores: Quality: 81.50 Length: 105
 Ratio: 1.509 Gaps: 4
Percent Similarity: 51.429 Percent Identity: 28.571

alignment_block:
us-10-048-197-2 x AC094388 ..

Align seg 1/1 to: AC094388 from: 1 to: 140984

```

5 His1leAryleuThr1leSerAlaLeuThra1leuValhrg1 21
|||||
47362 CACATATTGACAAAGAAAGTCCGTCATGTCGCAATGTTAACTGGCAGG 47411
21 Y.....CysValSerThrg1yAsnValAlaMetLySg1n.... 33
|||||
47412 GCATGGCAGGTGTCATCTGCGCTTCTCGCTCCGCCATAGACTAT 47461
34 .....AsnGlnThr1leGlnThr1leLeuLySg1yStrAsn 48
|||||
47462 TTGTTGAATTTCAGCCAGTCAGAGACCCATTGTCAAGGGGGAAGAAC 47511
49 LysGln.....
|||||
47512 TCCCAAGAACACGTTGTGTGGCCTCCATCTCCGCACACAGATGATTA 47561
51 .....Glu1leSerSerArgPheGly..... 57
|||||
47562 ACAAGCCACACTTTTGGGAGAAATTGTGACAGGTTTGGGGAATGTTAT 47611
58 ..SerAlaAspSer1leSerPheMet1leVal1Val1leLySphG1yAsn 73
|||||
47612 ATGCCCGACATGATGATGATGATGATGATGATGATGATGATGATGAT 47661
74 ThrAla1leLeuAla 78
|||||
47662 ACAAGCAATGTTGCC 47676

```

seq_name: gb_hlg:AL627314

seq_documentation_block:
LOCUS AL627314 196632 bp DNA linear HTG 30-JAN-2002
DEFINITION Mus musculus chromosome 4 clone RP23-354H24, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL627314
VERSION AL627314.5 GI:18152585
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus


```

misc_feature      fragment_chain:2"
59953..72074
/note="assembly_fragment:05167
fragment_chain:2"
72175..97636
/note="assembly_fragment:06788
fragment_chain:2"
97737..110968
/note="assembly_fragment:06944
fragment_chain:2"
111069..125967
/note="assembly_fragment:06767
fragment_chain:2"
126068..132726
/note="assembly_fragment:06915
fragment_chain:2"
132827..147689
/note="assembly_fragment:04978
fragment_chain:2"
147790..150330
/note="assembly_fragment:03806
fragment_chain:2"
150431..166175
/note="assembly_fragment:01355
fragment_chain:2"
166276..180026
/note="assembly_fragment:05619
fragment_chain:2"
180127..203936
/note="assembly_fragment:02077
fragment_chain:2"
204037..242827
/note="assembly_fragment:05860
fragment_chain:2"
242928..251725
/note="assembly_fragment:05353
fragment_chain:2"
251826..259837
/note="assembly_fragment:00313
fragment_chain:2"
clone_end:77
vector_side:right"
BASE COUNT      66786 a 63093 c 63992 g 64014 t 1952 others
ORIGIN

alignment_scores:
Quality:      81.50      Length:      77
Ratio:        1.598      Gaps:      1
Percent Similarity: 66.234      Percent Identity: 27.273

alignment_block:
US-10-048-197-2 x AL627253/rev ..
Align seg 1/1 to reverse of: AL627253 from: 1 to: 259837

31 LysGlnGlnAsnGlnGlnThrIleGlnGlnThrIleIleIleGlyLysTh 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148893 AAAAAAAAAAACCAACCTCCCTTCCTTCCTTCCTTCCTTCCTTC 148844
47 rAsnLysGlnGlnIleSerSerArgPheGlySerAlaAspSerIleSer 64
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148843 TATTCGCCCTCAGCTCTCTTCCTTCCTTCCTTCCTTCCTTCCTTC 148794
64 hemeIleValIleIleLysPheGlyHisThrAlaIleLeuAlaProAsn 80
||:|||||:|||||:|||||:|||||:|||||:|||||:
148793 TCTGTTTGTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 148744
81 ArgTrpGlnGlnIleLeuSerLeuIleIleSerPheLeuTrpValLysPr 97
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148743 .....TCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 148703
97 gTyraArgProLysAsnLeuSerPheTyLeu 107

```

```

148702 CTCCTCCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 148672
seq_name: gb_ba:AF169465
seq_documentation_block:
LOCUS      AF169465                267 bp      DNA      linear      BCT 09-AUG-2000
DEFINITION  Neisseria meningitidis strain Z2491 clone Em029 unknown sequence.
ACCESSION   AF169465
VERSION     AF169465.1 GI:9754673
KEYWORDS
SOURCE      Neisseria meningitidis.
ORGANISM    Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1 (bases 1 to 267)
AUTHORS    Perrin,A., Nassif,X. and Tinsley,C.R.
TITLE      Identification of regions of the chromosome of Neisseria
            meningitidis and Neisseria gonorrhoeae which are specific to
            pathogenic Neisseriae
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 267)
AUTHORS    Perrin,A., Nassif,X. and Tinsley,C.R.
TITLE      Direct Submission
JOURNAL    Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156
            rue de Vaugirard, Paris 75015, France
FEATURES
            Location/Qualifiers
            1..267
            /organism="Neisseria meningitidis"
            /strain="Z2491"
            /db_xref="taxon:487"
            /clone="Em029"
BASE COUNT      71 a 70 c 61 g 65 t
ORIGIN

alignment_scores:
Quality:      80.50      Length:      71
Ratio:        1.677      Gaps:      1
Percent Similarity: 67.606      Percent Identity: 33.803

alignment_block:
US-10-048-197-2 x AF169465 ..
Align seg 1/1 to: AF169465 from: 1 to: 267

7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyGlyAsp 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
14 CGCATCATCGTTTCGGCTGGCTTCGCTGGCTGGCTGGCTAGCAGGTTGCCG 63
23 lSerThrGlyAsnValAlaMetLysGlnGlnAsnGlnGlnThrIleGln 40
||:|||||:|||||:|||||:|||||:|||||:|||||:
64 CTCATCATTAATGTAACGTTTCGCACCAAGAACTTCAGGAA...CGTG 110
40 lnrThrIleIleLysGlyLysThrAsnLysGlnGlnIleSerSerArgPhe 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 CGCGGTTTGGCTTGGCGGTCACCAATGCCGTAATAATGACGAACGCGAGC 160
57 GlySerAlaAspSerIleSerPheMetIleValIleLysPheGlyHis 73
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
161 AATGAAGCATACGATCACTTACCCGACACTGTGGTAAGCGCGTGAC 210
73 sThrAlaIleLeu 77
||:|||||:|||||:|||||:|||||:|||||:|||||:
211 CAATGCTATGTAA 223

seq_name: gb_pat:A68918
seq_documentation_block:
LOCUS      A68918                267 bp      DNA      linear      PAT 06-MAY-1999
DEFINITION  Sequence 89 from Patent WO9802547.
ACCESSION   A68918
VERSION     A68918.1 GI:4759837

```



```

7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuValThrGlyCysVal 23
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
10 CGCATCAATCGTTTCGGCTGGCTTCGTTGGTGGCATAGCAGGTTGGG 59
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
23 LSerThrGlyAsnValAlaMetLeuGluGlnAsnGlnGlnThrIleGlu 40
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
60 CTCATCAATAAATGAATACGGTTTCGACCAAGCAACTCAGGACGGTGGC 100
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
40 LThrIleIleIleGlyGlyLysThrAsnLysGlnGlnIleLeuSerArgPhe 56
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
110 CGATTGCGCTGGCGCGACCCAAATGCCGTTAAATATGACCAACGGAGC 159
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
57 GlySerAlaAspSerIleSerPheMetIleValVal_IleLeuPheGlyH 73
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
160 AATGAAGAAGCATACGCTCAACTTTCACCGCACTGGGGTAAAGCCCGGAG 209
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
73 LThrIleAlaIleLeuAlaProAsnArgTyrGlnGlnIleIleLeuSerIle 89
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
210 CCATATCGTATGTACACAGTATATCAGACACATCGAGGCGTTTACCTCCG 259
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
90 LLeuSerPheLeuTyr 94
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
260 ATGCATATTTGTTGG 274

```

seq_documentation_block:	
LOCUS	38240 bp DNA linear INV 23-JAN-2002
DEFINITION	CeB0491
ACCESSION	CeB0491
	249307
	elegans cosmid B0491, complete sequence.

SOURCE ORGANISM
Collagen; diphterine synthasel; Ligand gated ionic channel; Sqr-1
Transfer RNA; tRNA-Ala.
Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota. *Nematode*. *Caenorhabditis elegans*.

Rhabditidae; Rhabditiidae; Peloderinae; Caenorhabditis
1 (sites)
none.

Genom

JOURNAL OF CLIMATE
MEDLINE
Science 282 (5396), 2012-2018 (1998)
99069613

2 (bases 1 to 38240)

TITLE Direct Submission

JOURNAL

COMMENT

Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jise@anger.ac.uk or tw@emulode.wustl.edu

Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.

Current sequence finishing criteria for the *C. elegans* genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is not the entire insert of clone B0491. It may be shorter because we only sequence overlapping sections, once, or longer because we arrange for a small overlap between neighbouring subclones.

tRNA

gene

tRNA

gene

gene

CDS

right end of clone B0491 is at 4602 in sequence Z48367.
The true left end of clone C33B4 is at 38141 in this sequence. The true right end of clone T24F1 is at 9741 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence Z49912.
The end of this sequence (38141..38240) overlaps with the start of sequence Z48367.
For a graphical representation of this sequence and its analysis name-B0491
name-B0491
http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
Location/Qualifiers

1. .38240

```
/organism="Caenorhabditis elegans"  
/db_xref="taxon:6239"
```

```
/chromosome="11"  
/clone="B0491"
```

```
complement(6683.6754)
/gene="B0491.t2"
```

```
/note="TGC Ala A-tRNA
predicted using tRNAscan-SE-1.11
```

preliminary predicti
similar to tRNA-Ala"

```
/product="tRNA-Ala"  
complement(6683..6754)
```

```

/gene="B0491.t2"
7178. .7249

```

```
/gene="B0491.t1"
/note="TGC Ala A-tRNA
```

predicted using tRNAscan-SE-1.11
preliminary prediction

```
similar to tRNA-Ala"
/product="tRNA-Ala"
```

```
7178. 7249
/gene="B0491.t1"
```

```
complement(join(9398. .10231,10279. .10419))
/gene="B0491.2"
```

```
complement(join(9398. .10231,10279. .10419))
/gene="B0491.2"
```

/note="cuticle collagen SQR-1, contains similarity to Pfam domain: PF01391 (Collagen triple helix repeat /20

(copies), Score=80.1, E-value=1.5e-20, N=2; PF01484 (Nematode cuticle collagen N-terminal domain) Score=33.1

E-value=2e-06, N=1
CDNA EST EMBL:Z14426 comes from this gene

CDNA EST	yk171d9.5	comes from this gene
CDNA EST	yk171d9.3	comes from this gene

CDNA	EST	yk303b12.3	comes from this gene
CDNA	EST	yk303b12.5	comes from this gene

```

/codon_start=1
/protein_id="CAA90084.1"

```

```
/db_xref="GI:3873801"
/db_xref="SWISS-PROT:P13114"
```

translation="MSVKLACYASVTATLMVCEFTMSTIYSEVDGFRKLDTEMN
EROSTNGIWKDIVITGSSKRPDPOVEET..."

KKTPTNGANGNPTCNCNADNKNKCPAGPSGPKGVPGVGLDGVPGVGADDIAPQ
ESYGCETCPBGWVGPBCATCBDBDCBTCTPBNHNPBPHADGSSAPBPQQPPAVPVFVNQ

GGGROGPDCTGGTGCCTCCDPCDCDAEACGCCAGCAGCGAGCCGCTGCGGCGG
GEPGRPGEEKRDAEHPIGRPRPKGPRGDQGPPTGPAQGNGLHGRPGEPCITVGPEGPSG

```
complement(join(11220. .11324,11372. .11872,12132. .12200,
2250 13253,\"
complement(join(11220. .11324,11372. .11872,12132. .12200,
```

```

22200: 122222)
22201: gene="B0491.3"
22202: element/40412/112222
22203: 122222 122222
22204: 122222 122222

```

```
complement(join(11220: .11872,12132. .12200,  
                12250. .12252))  
dense="PD0A01 3"
```

note="CDNA EST EMBL:M80153 comes from this gene

[illegible]

YK144d8.3 comes from this gene

gene
CDNA EST yk306d2.3 comes from this gene"
CDNA EST yk306d2.3 comes from this gene"
/codon_start=1
/protein_id="CA90085.1"
/db_xref="GI:3873802"
/translation="MNPEIDIDFGCTKRRODGYDLKRLKDGKRSVEVLMRLAOVTH
EKSACVPAQRASVTBGLKFAEAEVQDPMHFAKRLATVLEQATPEYMAETKRLKCC
SKKFEELDKLAKPRKPTAVLHLNGRKIKYVASITKLEKLAATPFOQPSHSYEER
NDFDLAAIKVNMKMENTFYVSKVTALVKDANAKVLTCEIDPEPSDAQEFADADA
KMLSKL"
complement(join(12961..13068,13118..13252,13294..13421,
13939..14074,14124..14454,14503..14622,14737..14849,
14896..14976,15027..15139,16101..16200))
/gene="B0491.4"
complement(join(12961..13068,13118..13252,13294..13421,
13939..14074,14124..14454,14503..14622,14737..14849,
14896..14976,15027..15139,16101..16200))
/gene="B0491.4"
/note="Weak similarity with the extracellular domains of
some ligand gated ionic channels
CDNA EST yk271a4.5 comes from this gene"
CDNA EST yk271a4.3 comes from this gene"
/codon_start=1
/protein_id="CA90086.1"
/db_xref="GI:3873803"
/translation="MSPLVIGCALLIOTSPROMOCRTKLTNRADGPHCFYVLL
LKEVEKSSVHSLSATPPNSDENITTEENLVKLYELVKSSIQFNVFGDITLNMKD
EKMKDKNGEENYVHEHLITFNSSALWTPHVADIDHTLCSGSGSYSDVDVDDGTIYA
RQFPLVASCVCYDVRKPFEEDDSCIFPTAEIPNEQPTLVLEGEKEMKRPVAAQK
ITEKRSQSLDIQVHSESPMVVVEKTSNLGLSLQWCVPAEKMSVRAALFE
PTLQAVLNASLPFLGDRTOIYVKLFSLHLOTCLQICPTPLNGFLGVPRLYIEF
VELITTSIMSLTILVMALSRVKNRPNPHRLSLKILNRVCCIEPERSAAYRR
YVEQTPESRRTSEVSOQGSVTRPAEDLYQDRRHITLANNMFSAIMSEVLEPMTFE
LEP"
complement(join(16538..16734,17233..17480,17538..17675,
17735..17951,18002..18122,18203..18285))
/gene="B0491.5"
complement(join(16538..16734,17233..17480,17538..17675,
17735..17951,18002..18122,18203..18285))
/gene="B0491.5"
/note="CDNA EST EMBL:M75776 comes from this gene; CDNA EST
yk56b6.5 comes from this gene
CDNA EST yk56b6.3 comes from this gene; CDNA EST yk121e5.5
comes from this gene
CDNA EST yk114g1.5 comes from this gene; CDNA EST
yk70e11.5 comes from this gene
CDNA EST yk187e10.3 comes from this gene; CDNA EST
yk187e10.5 comes from this gene
CDNA EST yk143h7.5 comes from this gene; CDNA EST
yk167a4.5 comes from this gene
CDNA EST yk501d3.3 comes from this gene; CDNA EST
yk375c6.3 comes from this gene
CDNA EST yk375c6.5 comes from this gene
CDNA EST yk341h7.3 comes from this gene
CDNA EST yk341h7.5 comes from this gene; CDNA EST
yk280d1.3 comes from this gene
CDNA EST yk280d1.5 comes from this gene; CDNA EST
yk292g4.3 comes from this gene
CDNA EST yk292g4.5 comes from this gene
CDNA EST yk292g4.5 comes from this gene; CDNA EST
yk262e9.3 comes from this gene
CDNA EST yk262e9.5 comes from this gene; CDNA EST
yk26b10.3 comes from this gene
CDNA EST yk236d10.5 comes from this gene; CDNA EST
yk22e2.3 comes from this gene
CDNA EST yk22e2.5 comes from this gene; CDNA EST
yk28e9.5 comes from this gene
CDNA EST yk290f5.5 comes from this gene; CDNA EST
yk329h3.5 comes from this gene
CDNA EST yk344g7.3 comes from this gene; CDNA EST
yk432a6.5 comes from this gene
CDNA EST yk452b11.5 comes from this gene; CDNA EST

```

      yk475b1.5 comes from this gene; CDNA EST
      cdna EST yk483f10.5 comes from this gene; CDNA EST
      yk350e11.3 comes from this gene; CDNA EST
      cdna EST yk573d2.3 comes from this gene; CDNA EST
      yk587h1.3 comes from this gene; CDNA EST
      cdna EST yk598g2.3 comes from this gene; CDNA EST
      yk64a1.3 comes from this gene; CDNA EST
      cdna EST yk497h9.5 comes from this gene; CDNA EST
      yk501d3.5 comes from this gene; CDNA EST
      cdna EST yk502a1.5 comes from this gene; CDNA EST
      yk550e11.5 comes from this gene; CDNA EST
      cdna EST yk573d2.5 comes from this gene; CDNA EST
      yk587h1.5 comes from this gene; CDNA EST
      cdna EST yk598g2.5 comes from this gene; CDNA EST
      yk64a1.5 comes from this gene; CDNA EST
/codon_start-1

alignment_scores:
    Quality:       79.00          Length:       102
    Ratio:         1.234          Gaps:        6
Percent Similarity: 62.745      Percent Identity: 30.392

alignment_block:
US-10-048-197-2 x CEB0491/rev ..

Align seg 1/1 to reverse of: CEB0491 from: 1 to: 38240

10 11eserAlaleuLeuthrAlaleuLeuVal1lThngIyCysValSerThrCl 26
   :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10632 TTAATCCAGCTTTTATTCGTCGTCTCGCTCGACACTGCTC..GTCGG 10586
      yasnVala1ametyLsgluGlnasnglInthrIleGluginFr...I 42
      ||||| |::::::::::: :|||||:|||||:|||||:|||||:
10585 TATGTAAAAAACAACAAAACAACTGAACAACTGCCGGAAGAATTTC 10536
      42 lellelysgLy..lystrhasnlysglngluileserSearpPhegly 57
      ::||| |::::::::::: ||| |:::::
10535 TTATMAAACGAGCAAGCATCCGGTAGCAAAATCGACAACTCTGCTTCGC 10486
      58 SerAlaspSerIleSerPheMetIleValIleIlyPheGlyHisTh 74
      ::::: |::::|
10485 AACACTAGGTGACTTACATTT..... 10465
      74 rAlalAleuAlaProAsnArgTrpGlnIleLeuSerIleIles 91
      ||| |::::: :| |::::| |::::| |::::| |::::|
10464 .....CTTGGCCGCAAT..TATTTGAATAATCCAAATTTGATTAATTT 10425
      91 erPheLeutrIlyAllysProTyrrArgProIylAsnaseSerPheIreu 107
      || |::::| |::::| |::::| |::::| |::::|
10424 CAGAGATGCTGTAAHA.....CTTGCGTGTATATG 10393
      108 ThrAla 109
      |||||
10392 ACGGCT 10387

seq_name: gb_ba:NMA222491

seq_documentation_block:
LOCUS NMA222491 349061 bp DNA linear BCT 04-DEC-2000
DEFINITION Neisseria meningitidis serogroup A strain Z2491 complete genome;
segment 2/7.
ACCESSION AL162753 AL157959
VERSION AL162753.2 GI:7379120
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis Z2491.
Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
1 (bases 1 to 349061)
Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
Davies,R.M., Davis,P., Devlin,K., Felwell,T., Hamlin,N.,
Holroyd,S., Jagsels,K., Leather,S., Moule,S., Mungall,K.,

```

```

TITLE
    Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
    Skelton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.
    Complete DNA sequence of a serogroup A strain of Neisseria
    meningitidis 22491
JOURNAL
    Nature 404 (6777), 502-506 (2000)
MEDLINE
    20222556
REFERENCE
    2 (bases 1 to 349061)
AUTHORS
    Parkhill,J.
TITLE
    Direct Submission
JOURNAL
    Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
    sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
    Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
    Notes:
    Details of N. meningitidis sequencing at the Sanger Centre are
    available on the World Wide Web.
    (URL, http://www.sanger.ac.uk/Projects/N\_meningitidis/).
FEATURES
    source
        1. 349061
        /organism="Neisseria meningitidis 22491"
        /strain="22491"
        /db_xref="taxon:122587"
        /note="serogroup: A"
        /complement(24..206)
        /note="ATR repeat; hmms hit to HMM ATR (1 - 183), score:
        310.39"
        /label=ATR
        209..212
        220..681
        /gene="NMA0368"
        220..681
        /gene="NMA0368"
        /note="NMA0368"
        /note="NMA0368, probable integral membrane protein, len:
        153 aa; contains four probable transmembrane domains"
        /codon_start=1
        /transl_table=11
        /product="putative integral membrane protein"
        /protein_id="CAB83669.1"
        /db_xref="GI:7379121"
        /db_xref="SPTREMBL:O9JWH7"
        /translation="MOEONRKSPPPIVNLVSLVALMSLSNVAFTLGNGSMGLTV
        LILGSPASIDIRYCAVANYVWLAIVLALRRKVPVHAAFGLALVFSYVAVY
        DRAGNLSDIYRGAGFYLMYAAFAVAVSIGTPAGNKKERKASADGTRKNV"
        671..1492
        /gene="hemk"
        671..1492
        /gene="hemk"
        /note="NMA0369, hemk, Hemk protein, len: 273 aa; similar
        to e.g. HEMK_ECOLI:P37186 HEMK protein (277 aa), fasta
        scores: E(): 0, 42.38 identity in 279 aa overlap. Contains
        PS00092 N-6 Adenine-specific DNA methylases signature"
        /codon_start=1
        /transl_table=11
        /product="Hemk protein"
        /protein_id="CAB83670.1"
        /db_xref="GI:7379122"
        /db_xref="SPTREMBL:O9JWH6"
        /translation="MTFDKWLGSKLPRKNEARMLQYVSEYTRVQLTFRGSEMPDEV
        RQRDRLAQRRLNGEVPAYILGARREYGRFTYVPSVLTPRETEHIVAVARLPEN
        GRWVDLTGSGAVALYVALERPDPAFVASDISPPALERTKKAADLGAARVAVARLAPN
        DTMPSEGGKDIIVSNPIYIENGDKILSGDRLREFQIATLPFSGLSCTIRLACGASF
        DRLAEGGFLLLEHGFQGAAGVLAENFSGVETLPDLAGIDRVTLGRYKMKHLK"
        831..840
        /gene="hemk"
        /note="hemk"
        /note="Core DNA uptake sequence: gccgctcgaa"
        /label=DUS
        1162..1171
        /gene="hemk"
        /note="Core DNA uptake sequence: gccgctcgaa"
        /label=DUS
        1187..1207
        /gene="hemk"
        /note="PS00092 N-6 Adenine-specific DNA methylases
        signature"

misc_feature
    complement(1279..1288)
    /note="Core DNA uptake sequence: gccgctcgaa"
    /label=DUS
    1565..1568
    1578..1596
    /gene="NMA0370"
    1578..1596
    /gene="NMA0370"
    /note="NMA0370"
    462 aa; similar to hypothetical proteins e.g. Y335_HAEIN
    P44640 hypothetical protein H10325 (450 aa), fasta scores:
    E(): 0, 49.88 identity in 464 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="putative integral membrane protein"
    /protein_id="CAB83671.1"
    /db_xref="GI:7379123"
    /db_xref="SPTREMBL:O9JWH5"
    /translation="MNAVVAIVVMLVLSIRVHVLTITGAFVGGAVACPLONIA
    DAAGVSOAGIIPVENKGLGEGAKIALSVAMGAPRMAITHSGLPQOLAGAVRKLNR
    RLTACVTRGELVTVTFWFLPYGRATPLNELILGNHSAPQDLVDKNINVMAMAIPLR
    GMLAGLILATVHRKFRLLQSNNAIDTGNADANRPSAYSRSIAAVALAVCAFIOL
    MYEDSLVGLAMLGFAVPMMLGVINRDKADVGEIGIKMAMVGFIMIAOGFAAVMA
    TGHIOPLVSSNAIFGNSKGMALMLVGLVITWGIGSSSEFTPLTITATVPLCYGL
    GFSPLATVAIVGTAGALDGAGSPASDSTLGTMTGLNADGQHDHIDSVTFPIHNTIP
    2967..3191
    2967..3191
    /gene="slyx"
    2967..3191
    /gene="slyx"
    /note="slyx"
    /note="NMA0371, slyx, SLIX protein homolog, len: 74 aa;
    similar to SLIX_HAEIN P44759 SLIX protein homolog (73 aa),
    fasta scores: E(): 0.0017, 33.88 identity in 74 aa
    overlap, and SLIX_ECOLI:P30857 SLIX protein (72 aa), fasta
    scores: E(): 0.24, 32.48 identity in 68 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="slyx protein homolog"
    /protein_id="CAB83672.1"
    /db_xref="GI:7379124"
    /db_xref="SPTREMBL:O9JWH4"
    /translation="MDAVOEERHITLEIQLSALQEDVYIAGLNAVAELRQTLDDLOQA
    OLRLYOKMODRNPDAQEPYSRLRDEIRPHF"
    complement(3271..3423)
    /gene="NMA0372"
    complement(3271..3423)
    /gene="NMA0372"
    /note="NMA0372, unknown, questionable CDS, len: 50 aa"
    /codon_start=1
    /transl_table=11
    /product="very hypothetical protein NMA0372"
    /protein_id="CAB83673.1"
    /db_xref="GI:7379125"
    /db_xref="SPTREMBL:O9JWH3"
    /translation="MFGRLSPGSDGFLLAVYEYPVGICLOGKTAIRYRSGGKRPHRI
    KGLIS"
    3562..3571
    /note="Core DNA uptake sequence: gccgctcgaa"
    /label=DUS
    complement(3586..4356)
    /gene="thiF"
    complement(3586..4356)
    /gene="thiF"
    /note="NMA0373, thiF, probable ThiF protein, len: 256 aa;
    similar to e.g. THIF_ECOLI:P30138 THIF protein (251 aa),
    fasta scores: E(): 0, 43.18 identity in 246 aa overlap,
    and MOEB_ECOLI:P12282 molybdopterin biosynthesis MOEB
    protein. (249 aa), fasta scores: E(): 0, 43.98 identity in
    244 aa overlap (note that N.m. does not have orthologs of
    any other molybdopterin biosynthesis proteins). Contains
    Pfam match to entry PF00899 ThiF_family, ThiF family"
    /codon_start=1

```



```
*
* 3018 3117: gap of 100 bp
* 3118 3807: contig of 690 bp in length
* 3808 3907: gap of 100 bp
* 3908 4576: contig of 669 bp in length
* 4577 4676: gap of 100 bp
* 4677 5353: contig of 677 bp in length
* 5354 5453: gap of 100 bp
* 5454 6134: contig of 681 bp in length
* 6135 6234: gap of 100 bp
* 6235 6904: contig of 670 bp in length
* 6905 7004: gap of 100 bp
* 7005 7690: contig of 686 bp in length
* 7691 7790: gap of 100 bp
* 7791 8484: contig of 694 bp in length
* 8485 8584: gap of 100 bp
* 8585 9275: contig of 691 bp in length
* 9276 9375: gap of 100 bp
* 9376 10095: contig of 720 bp in length
* 10096 10195: gap of 100 bp
* 10196 10912: contig of 717 bp in length
* 10913 11012: gap of 100 bp
* 11013 11713: contig of 701 bp in length
* 11714 11813: gap of 100 bp
* 11814 12507: contig of 694 bp in length
* 12508 12607: gap of 100 bp
* 12608 13304: contig of 697 bp in length
* 13305 13404: gap of 100 bp
* 13405 14076: contig of 672 bp in length
* 14077 14176: gap of 100 bp
* 14177 14866: contig of 690 bp in length
* 14867 14966: gap of 100 bp
* 14967 15644: contig of 678 bp in length
* 15645 15744: gap of 100 bp
* 15745 16418: contig of 674 bp in length
* 16419 16518: gap of 100 bp
* 16519 17209: contig of 691 bp in length
* 17210 17309: gap of 100 bp
* 17310 18000: contig of 691 bp in length
* 18001 18100: gap of 100 bp
* 18101 18794: contig of 694 bp in length
* 18795 18894: gap of 100 bp
* 18895 19595: contig of 701 bp in length
* 19596 19695: gap of 100 bp
* 19696 20352: contig of 657 bp in length
* 20353 20452: gap of 100 bp
* 20453 21143: contig of 691 bp in length
* 21144 21243: gap of 100 bp
* 21244 21907: contig of 664 bp in length
* 21908 22007: gap of 100 bp
* 22008 22700: contig of 693 bp in length
* 22701 22800: gap of 100 bp
* 22801 23489: contig of 689 bp in length
* 23490 23589: gap of 100 bp
* 23590 24279: contig of 690 bp in length
* 24280 24379: gap of 100 bp
* 24380 25063: contig of 684 bp in length
* 25064 25163: gap of 100 bp
* 25164 25862: contig of 699 bp in length
* 25863 25962: gap of 100 bp
* 25963 26655: contig of 693 bp in length
* 26656 26755: gap of 100 bp
* 26756 27442: contig of 687 bp in length
* 27443 27542: gap of 100 bp
* 27543 28228: contig of 686 bp in length
* 28229 28328: gap of 100 bp
* 28329 29018: contig of 690 bp in length
* 29019 29118: gap of 100 bp
* 29119 29808: contig of 690 bp in length
* 29809 29908: gap of 100 bp
* 29909 30602: contig of 694 bp in length
* 30603 30702: gap of 100 bp
* 30703 31398: contig of 696 bp in length
* 31399 31498: gap of 100 bp
*
* 31499 32195: contig of 697 bp in length
* 32196 32295: gap of 100 bp
* 32296 33012: contig of 717 bp in length
* 33013 33112: gap of 100 bp
* 33113 33824: contig of 712 bp in length
* 33825 33924: gap of 100 bp
* 33925 34613: contig of 689 bp in length
* 34614 34713: gap of 100 bp
* 34714 35402: contig of 689 bp in length
* 35403 35502: gap of 100 bp
* 35503 36190: contig of 688 bp in length
* 36191 36290: gap of 100 bp
* 36291 36951: contig of 661 bp in length
* 36952 37051: gap of 100 bp
* 37052 37741: contig of 690 bp in length
* 37742 37841: gap of 100 bp
* 37842 38530: contig of 689 bp in length
* 38531 38630: gap of 100 bp
* 38631 39357: contig of 727 bp in length
* 39358 39457: gap of 100 bp
* 39458 40125: contig of 668 bp in length
* 40126 40225: gap of 100 bp
* 40226 40920: contig of 695 bp in length
* 40921 41020: gap of 100 bp
* 41021 41714: contig of 694 bp in length
* 41715 41814: gap of 100 bp
* 41815 42519: contig of 705 bp in length
* 42520 42619: gap of 100 bp
* 42620 43305: contig of 686 bp in length
* 43306 43405: gap of 100 bp
* 43406 44103: contig of 698 bp in length
* 44104 44203: gap of 100 bp
* 44204 44876: contig of 673 bp in length
* 44877 44976: gap of 100 bp
* 44977 45644: contig of 668 bp in length
* 45645 45744: gap of 100 bp
* 45745 46429: contig of 685 bp in length
* 46430 46529: gap of 100 bp
* 46530 47219: contig of 690 bp in length
* 47220 47319: gap of 100 bp
* 47320 47994: contig of 675 bp in length
* 47995 48094: gap of 100 bp
* 48095 48798: contig of 704 bp in length
* 48799 48898: gap of 100 bp
* 48899 49595: contig of 697 bp in length
* 49596 49695: gap of 100 bp
* 49696 50400: contig of 705 bp in length
* 50401 50500: gap of 100 bp
* 50501 51167: contig of 667 bp in length
* 51168 51267: gap of 100 bp
* 51268 51970: contig of 703 bp in length
* 51971 52070: gap of 100 bp
* 52071 52736: contig of 666 bp in length
* 52737 52836: gap of 100 bp
* 52837 53511: contig of 675 bp in length
* 53512 53611: gap of 100 bp
* 53612 54295: contig of 684 bp in length
* 54296 54395: gap of 100 bp
*
* alignment_scores:
*      Quality: 78-50
*      Ratio: 1.670
*      Percent Similarity: 47.475
*      Length: 99
*      Gaps: 3
*      Percent Identity: 27.273
*
* alignment_block:
* US-10-048-197-2 x AC102510
*
* Align seg 1/1 to: AC102510 from: 1 to: 65312
*
* 8 leuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrIleCysValse 24
* |||:|||||
* 2340 CTGGTCATCTTCACTGTCACACTGTTGTACTAGTACGACAAATGTAAAC 2389
```



```

24 trhrghysanvalalameltygluclnmsngclntrhrleuglnt 41
2390 ACCGTGACAT.....ACGATGAACGACG 2412
41 hrlellelysglythrnsanlysglclulleserfargphgly 57
2413 ATGTAGTCACGCAAA..... 2429
58 SerialabpserliserPhemetilevallelyspheglyhsth 74
2430 .....GCTGTCATAGTGAAGCCACATC 2452
74 lalalleleualaproasnartrpslnluleuserlelleles 91
2453 ACCGTTCAAGATGTTTCACGTTGCAGTAATCAACCCCAACGACGATATT 2502
91 erPhelaurpvallyspotyargprolyasaneuserphetyr 106
2503 CAGTGTGTTTCG.....CAAGTAATTAACCTTAATTCAT 2537
seq_name: gb_pat:BD003755
seq_documentation_block:
LOCUS BD003755 8136 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003755
VERSION BD003755.1 GI:18631716
KEYWORDS JP 2001501833-A/75.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 8136)
AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bala,S.C.,
TITLE Fannon,M. and Dougherty,B.A.
JOURNAL Polynucleotide of Streptococcus pneumoniae and sequence
COMMENT Patent: JP 2001501833-A 75 13-FEB-2001;
HUMAN GENOME SCIENCES INC
OS Unidentified
PN JP 2001501833-A/75
PD 13-FEB-2001
PE 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029860
PI CHARLES A KUNSCH,GIL H CHOI,PATRICK J DILLON,CRAIG A ROSEN, PI
STEVEN C BARASH,
PC MICHAEL FANNON,BRIAN A DOUGHERTY
PC C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19,
PC C12N1/21,
PC C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC
G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH key Location/Qualifiers
FT source 1. 8136
/organism='Unidentified',
FEATURES
source Location/Qualifiers
1. 8136
/organism='unidentified'
/db_xref='taxon:32644'
BASE COUNT 2249 a 1481 c 1983 g 2423 t
ORIGIN
alignment_scores:
Quality: 78.00 Length: 127
Ratio: 1.200 Gaps: 6
Percent Similarity: 51.181 Percent Identity: 28.346
alignment_block:
US-10-048-197-2 x BD003755 ..
Align seg 1/1 to: BD003755 from: 1 to: 8136
10 lleserialleuleuthrala...leuleualtrhglycysvalserth 25

```

1926 CTGAGGACCTTGGTACAGCTGTGATTCCTGTAACGGCTCTGTTCTACT 1975
25 T.....GlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleSer 40
1976 CATTTGGAAAATGTCACGACAGATTTTGGATCCGACACAGCATATGATG 2025
40 InThrIleLe..... 43
2026 AGCGGATCTCTCGTTAGGAATATTGCGATTATCATATCATCTGTTAGCG 2075
44 LysGlyLysThrAsnLysGlnIleSerSer 55
2076 AGTCGTGCTGTTGGTGAAGGAGAAAGACAAAGATGATTAATCTGAGTCT 2125
55 gphegJserAlaAspSerIleSerPhe.....MetIle 67
2126 GCATTTTTCGGAAGATACGCTAGCGTGGGTAGCTGTTATCCTGATGCGCA 2175
67 aValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 83
2176 TTGTTCTTGATTTACGAGCTGGATATPACTCTGATGCT..... 2213
84 GluIleLeuSerLeuIleLeuSerPheLeuTyrPvalLysProTyrArgPr 100
2214 ...CTTTGTCCCTGTGCATTTCTTTCTTATTCCTTCAAAAGCCCTTCC 2260
100 oLysAsnLeuSerPheTyrLeuThrAlaLys 110
2261 ACGT.....TTTGGTCTACACTCAAG 2282
seq_name: gb_ba:AE008532
seq_documentation_block:
LOCUS AE008532 10393 bp DNA linear BCT 13-SEP-2001
DEFINITION Streptococcus pneumoniae R6 section 148 of 184 of the complete
genome
VERSION AE008532 AE007317
KEYWORDS AE008532.1 GI:15459334
SOURCE
ORGANISM Streptococcus pneumoniae R6.
Streptococcus pneumoniae R6
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 10393)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Khoja,H., Kraft,A., Lagace,R.,
Leblanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McHenry,S., McHenry,M., Mcleaster,R., Mundy,C., Nickes,T.I.,
Norris,F.H., O'gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
TITLE Genome of the Bacterium Streptococcus pneumoniae Strain R6
J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10393)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., Lagace,R.,
Leblanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McHenry,S., McHenry,M., Mcleaster,R., Mundy,C., Nickes,T.I.,
Norris,F.H., O'gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
TITLE Direct Submission
J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10393)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., Lagace,R.,
Leblanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McHenry,S., McHenry,M., Mcleaster,R., Mundy,C., Nickes,T.I.,
Norris,F.H., O'gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
TITLE Direct Submission
J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10393)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., Lagace,R.,
Leblanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McHenry,S., McHenry,M., Mcleaster,R., Mundy,C., Nickes,T.I.,
Norris,F.H., O'gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
TITLE Direct Submission
J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10393)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., Lagace,R.,
Leblanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McHenry,S., McHenry,M., Mcleaster,R., Mundy,C., Nickes,T.I.,
Norris,F.H., O'gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
TITLE Direct Submission
J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10393)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., Lagace,R.,
Leblanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McHenry,S., McHenry,M., Mcleaster,R., Mundy,C., Nickes,T.I.,
Norris,F.H., O'gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
TITLE Direct Submission
J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10393)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., Lagace,R.,
Leblanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McHenry,S., McHenry,M., Mcleaster,R., Mundy,C., Nickes,T.I.,
Norris,F.H., O'gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
TITLE Direct Submission
J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10393)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., Lagace,R.,
Leblanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McHenry,S., McHenry,M., Mcleaster,R., Mundy,C., Nickes,T.I.,
Norris,F.H., O'gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
TITLE Direct Submission
J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10393)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., Lagace,R.,
Leblanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McHenry,S., McHenry,M., Mcleaster,R., Mundy,C., Nickes,T.I.,
Norris,F.H., O'gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
TITLE Direct Submission
J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10393)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., Lagace,R.,
Leblanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McHenry,S., McHenry,M., Mcleaster,R., Mundy,C., Nickes,T.I.,
Norris,F.H., O'gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
TITLE Direct Submission
J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10393)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., Lagace,R.,
Leblanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McHenry,S., McHenry,M., Mcleaster,R., Mundy,C., Nickes,T.I.,
Norris,F.H., O'gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
TITLE Direct Submission
J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10393)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Hann,A., Khoja,H.,

```

/organism="Streptococcus pneumoniae R6"
/db_xref="taxon:171101"
102..683
/gene="xpt"
/note="spr1662"
102..683
/gene="xpt"
/EC_number="2.4.2.-"
/note="this sequence is longer than the other published
sequence"
/codon_start=1
/transl_table=11
/label_spr1662
/product="Xanthine phosphoribosyltransferase"
/db_xref="GI:15459335"
/translation="MKLEERTLKDGHIIDNLIKVDSEFLTHQVDESLMREIKGVAE
KFAATGTRKVTIEASGIAPAVFTAEALNPVIMPAKKAKNITNNEGILTAQVSEFTKO
VSTVSIAGKFLSPEDKVLIDDFLANGAARKGLIOLIEAGATVOAGIYIEKSFOD
GHDLEKAGYPLSIALRDRFENGQVFEKADL"
683..1945
/gene="pbux"
/note="spr1663"
683..1945
/gene="pbux"
/codon_start=1
/transl_table=11
/label_spr1663
/product="Nucleobase:cation symporter for xanthine"
/protein_id="AAL00466.1"
/db_xref="GI:15459336"
/translation="MQTOEKHSQAAYLGLOHLAMYSGLIVPIMTALGSAEOLT
YLITDIFMGCVAATFLOLNKRYFGIGLPVVLVAFQSVAPLIMIGSHSGAMFAL
IASGITVVLVSGIESKVAANLEPISIVTSGVITIGLIPVALIGNMGNNDPEPGOSIL
LAATVVLILILNIFTKGFTKSIILGLVYGTAIAATMGLVDFSPVAVAPVHYVTP
LYFGMPFEISISVAMCIIATVSMVESTGYLALSDTMDPDSITPLNRGYSREGIAY
LLGGIFNTPYTFGSONVGLVKSIGIKRPIIYAAGFVLVLELFRFGALQIIPSS
VLGGAMLVMEGFVSIQGMOTIARVDRNNHNLIAVSTIAGVGLNNSLDFVSMPTA
FOMFENGIVVSSILATVLAVALNHKKK"
complement(11996..2457)
/gene="dpnd"
/note="spr1664"
complement(11996..2457)
/gene="dpnd"
/EC_number="2.1.1.-"
/codon_start=1
/transl_table=11
/label_spr1664
/product="Restriction system of S. pneumoniae"
/protein_id="AAL00467.1"
/db_xref="GI:15459337"
/translation="MKTQOLVAEEVYDLKYTPDYETESRYDNLISLIVTCLSDPDC
VRLSENKMGDEKOLAMKEKYGWEGKIDPEWLIHSHHRLILIGELLESRLKKY
TVETITLSRIYSIENPDEAERLVREKYSCEIVLADDFQDYDTIYE"
complement(2454..3218)
/gene="dpnc"
/note="spr1665"
complement(2454..3218)
/gene="dpnc"
/EC_number="3.1.21.4"
/codon_start=1
/transl_table=11
/label_spr1665
/product="Type II restriction enzyme DpnI (dpnc)"
/protein_id="AAL00468.1"
/db_xref="GI:15459338"
/translation="MELFNLIELVETYSNQRRIITEDVYRQSYCPNCGNNPLNH
FENNRPVADFCNHCSEFEELSKKGNFSSTINDGAVATMNRKRVQADNNNPFELIYV
KNFVNPLVLEKQGLVPKPSIIORRPLAPFARAGWIGCNIDLSQVSKGRIFLYODG
QVRDEKATKEFKQGLFLKSSLSRGMTEILNCIDKIGSSEFTLEDMMRFESDLKN
IFVNNHAIKEIRIOLIRDKETIEFKGNGKVKKL"
complement(3399..3824)
gene,

```

```

CDS
/gene="spr1666"
complement(3399..3824)
/gene="spr1666"
/note="Similar to E. coli PaaI protein involved in
phenylacetic acid degradation"
/codon_start=1
/transl_table=11
/label_spr1666
/product="Conserved hypothetical protein"
/protein_id="AAL00469.1"
/db_xref="GI:15459339"
/translation="MKYQETMKDFHDAISAFENYIEKMDGHVVTYVNSLN
YGNAGGYLFTICDQISGLVVISIGDVTQSSINLAKKGLDVLITNGECVHQG
RTQVMDVDITNOEGERNCKAKETPFYVGQRSEERVSI"
complement(3873..5354)
/gene="galT"
/note="spr1667"
complement(3873..5354)
/gene="galT"
/EC_number="2.7.7.10"
/codon_start=1
/transl_table=11
/label_spr1667
/product="galactose-1-phosphate uridylyltransferase"
/protein_id="AAL00470.1"
/db_xref="GI:15459340"
/translation="MTLVKRFVTHVISESFEEMDRITLNRVLSRYEGEVLFEVNTL
DKLIDKQOLVEEYVRLTIEDSOTARVPIGTLEMDLITPCSPQVNDPFEATAYSE
QAIEDFYLOSKNDYIKLKATAKINAVRPSPDVGELITITLSKEPDPEAIYAEV
QASNPQOCOLTENGYGRVNHPRARSNRIIREMVGQMGFOYSYAEVFNHCTEL
DGHAPKMAISROSREERLAIYEOFPGTFAGSNADLPYVGSILTHDHVOGGRVFPME
LAEEDGPHHTTIPPIAKRDGQELDLVLDNOTSAEHPDGIYHPKDVQOHIKKNI
ALIVDSGALIPRLKEVEEDVASYLGEAVTVADYHOEWADOLKSOHPLDPKEAL
complement(5373..6551)
/gene="galK"
/note="spr1668"
complement(5373..6551)
/gene="galK"
/EC_number="2.7.1.6"
/codon_start=1
/transl_table=11
/label_spr1668
/product="galactokinase"
/protein_id="AAL00471.1"
/db_xref="GI:15459341"
/translation="WTQHLTAETLRKDFLAVGEQVDQTFSPGRINLIGEHYDNG
HFFPVALSIGTGAARKRDQVLYRFSANFEDKGLIEVLADLKEKEHNTNPKGV
LHFLQENGHYIDKGFDFEYVYGNIPNGSGSSASLEILGVAAEHLFDLKIERLDIYK
IGRQETENNFGVNSGINDQFAIGMGADORAITYLDNTLEVDLPDLKDNYVLTNTN
KRRELDASKINERARCEKAEVETQVALDIOTIGTDEDAVQYIYLKDKENLRAR
HAYLENDRTLKQAALQAGDLEFFRGLMAASHVSLHEDYEVLELDLTVHPMAQEG
VLGARITGAGFSCALVALYOKTIVEAEKAVGRHYEEVAPSPSYIAEVAGTRVLD
6715..7722
/gene="galR"
/note="spr1669"
complement(6715..7722)
/gene="galR"
/codon_start=1
/transl_table=11
/label_spr1669
/product="GalR, member of GalR-Laci family of
transcriptional regulators, binds DNA; regulator of gal
operon"
/protein_id="AAL00472.1"
/db_xref="GI:15459342"
/translation="WATLKIDIAOLASVIATSVRYLNKDSLSVTEETRRRLITVAE
IGYTHLKTGSHKPKOKIAITOWVSEGEIDDLYYOIRGIEKRAOEIDYILARF
NDHPFTLSEEVIGLICIGKFSRAOISAEEOYKPLVDFNNTYSIGTCTITPTPYAM
KQVVDYPLSQGMDRIGITGLGETTDOEITIODKRLNPNYSQARGITHDELVFGGR
FTAGSGYDMKEAIOISLGDOLPPAPFPAASDSIAIGALRALQEGISLIPRVSLSFND

```

```

TSLTKOVYPPLSITVYTEEMGRAGMIDILNKEVLHGRIKPSITLMLGTRLTRESTLNG
E"
gene      complement(7768. .8805)
          /gene="adhB"
          /note="SPI1670"
          /complement(7768. .8805)
CDS
          /gene="adhB"
          /EC.number="1.1.1.1"
          /note="putative"
          /codon_start=1
          /transl_table=11
          /label="SPI1670"
          /product="Alcohol dehydrogenase"
          /protein_id="AA100473.1"
          /db_xref="GI:15459343"

alignment_scores:
  Quality: 78.00      Length: 127
  Ratio: 1.200        Gaps: 6
  Percent Similarity: 51.181      Percent Identity: 28.346

alignment_block:
  US-10-048-197-2 x AE008532/rev ..

Align seg 1/1 to reverse of: AE008532 from: 1 to: 10393

10 115SerAlaLeuLeuThrAla...LeuLeuValThrGlyCysValSerTh 25
   ::::|||||:::|||||:::|||||:::|||||:::|||||:::
10071 CTAGGAGCCTTGTGTAACAGCTGTGATCTCTGTAACGGGCTCTGTCTAGT 10022
      :
25  T.....GlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
   :|||||:::|||||:::|||||:::|||||:::|||||:::
10021 CATTTGGAAATGTCACGACAGATTTTGGCATCCGCCAACGACGATGATG 9972
      :
40 1nThrIleIle.....
   :|:::|
9971 AGGGGATTCCTGCTAGGAAATTATGGCATTAATACTCATCTGTTAGCG 9922
      :
44  ....LysGlyLysThrAsnLysGlnGluIleSerSerar 55
   |||||||:::|||||:::|||||:::|||||:::|||||:::
9921 AGTCTGTGCTGTGTAAGGAAAGACAAAGAAATAGATCTATCTGAGTCT 9872
      :
55  gPheGlySerAlaAspSerIleSerPhe.....MetIleV 67
   |||:::|||||:::|||||:::|||||:::|||||:::
9871 GCATTTTCTGGAAGATACCTAGGAGTGGTAGCTGTTATCTGATGCCGA 9822
      :
67  a1ValIleLeuSerPheGlyHisThrAlaIleLeuAlaProAsnArgTyrGln 83
   :|||||:::|||||:::|||||:::|||||:::|||||:::
9821 TTCTTCTCTCATTTACGGACTGGTATATCTCATGATCCT..... 9784
      :
84  GluIleLeuSerLeuIleIleSerPheLeuTyrValLysProTyrArgPr 100
   :|||||:::|||||:::|||||:::|||||:::|||||:::
9783 ...CTTTTGCCTTGTCAATTTCTTTCTTATCTTCAAAAGCCCTTCC 9737
      :
100  oLysAsnLeuSerPheTyrLeuThrAlaLys 110
   |||:::|||||:::|||||:::|||||:::|||||:::
9736 ACCT.....TTTGGTCTACACTCAAG 9715

seq_name: gb_ba:AE007477

seq documentation_block:
LOCUS      AE007477          12841 bp      DNA          linear      BCT 31-AUG-2001
DEFINITION Streptococcus pneumoniae TIGR4 section 160 of 194 of the complete
            genome.
ACCESSION  AE007477 AE005672
VERSION    AE007477.1 GI:14973352
KEYWORDS
SOURCE
ORGANISM   Streptococcus pneumoniae TIGR4.
            Streptococcus pneumoniae TIGR4
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Streptococcus.
REFERENCE  1 (bases 1 to 12841)
AUTHORS   Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,

```

```

TITLE
JOURNAL    Complete genome sequence of a virulent isolate of Streptococcus
MEDLINE    pneumoniae. 293 (5529), 498-506 (2001)
PUBMED     21357209
REFERENCE  11463916
AUTHORS    2 (bases 1 to 12841)
            Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
            Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
            Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,M.C., Peterson,J.D.,
            Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D.,
            Holtzapple,E., Knouri,H., Wolf,A.M., Uterback,T.R., Hansen,C.L.,
            McDonald,L.A., Feldblyum,T.V., Anguillo,S., Dickinson,T.,
            Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O.,
            Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and
            Fraser,C.M.

FEATURES
  source
    organism="Streptococcus pneumoniae TIGR4"
    strain="TIGR4"
    /db_xref="taxon:170187"
    complement(121. .525)
    /gene="SPI151"
    complement(121. .525)
    /gene="SPI151"
    /note="identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AAK75923.1"
    /db_xref="GI:14973353"
    /translacion="MKDFPDAISAFENYEIEKMDGHVVVTKVNSLWYGNAG
    GYLTFDCOISGLVVISGLDGVLTSSINVLKAGKLDVLTIKECEVHQGRTCYMD
    VDIRNOEGRNCKATFTMTFTGQREERVRH"
    complement(596. .2077)
    /gene="SPI152"
    complement(596. .2077)
    /gene="SPI152"
    complement(596. .2077)
    /note="identified by match to PFAM protein family HMM
    PF01087"
    /codon_start=1
    /transl_table=11
    /product="galactose-1-phosphate uridylyltransferase"
    /protein_id="AAK75924.1"
    /db_xref="GI:14973354"
    /translacion="MTLVNKFYTHVYISSEPEMDRITYLTNRVLARVGEVLEVTNL
    DKLIDKQLYBEAVRLETIEDSOTARILILAEMLDVTPCPSQVNRPFATYASNP
    QATIEDFYLSQKNDITKLKATARNATVVPDSYGLLETTINLSKEPKETVAKVL
    QASNPQOCOLCLENEGYGRVNHPARSNHRIIRFEMVQEMGFQSPAYRNEHIFL
    DGOHRPNAISROSEPERLAIVDQPGYPAGSNADLPVIGSGLTIDHYOGGRHVFPE
    LAPLDKARFAGFEQVAGIYKMPWSVLRITSDSKEDLINLADKLQEMROYSDPAVO
    ILAETDRPHHTITPPIARKRDGQFELDIVLRNONSAYRPGIYRPHKDVQHIKKEI
    GLIEVGLAIIIPRIKREVEQVAYSLVEAATYVADYHOBMDQLKSNPDLTRKRAL
    AIVKSVGAIFARVLEDGVTKQEGQATFRKEVEQGLLD"
    complement(2096. .3274)
    /gene="SPI153"
    complement(2096. .3274)
    /gene="SPI153"
    complement(2096. .3274)
    /note="identified by match to TIGR protein family HMM
    TIGR00131"
    /codon_start=1
    /transl_table=11
    /product="galactokinase"
    /protein_id="AAK75925.1"

```

```

/db_xref="GI:14973355"
/translation="MAOHITPALKRKDLAVFGADQTFSPGRINLIGEHDTYNGG
HVEPAISICTGARKRDVLRFTYSANFEDKGIIEVPLADLKEFEHMTNTYPKCV
LHPIOEAGHYIDKGFDFYVYVINGAGLSSASLELTGVAAEHLFDLKERDLVK
IKGTJENNTIGNVSGIMDOFAIGMGADRAIYLDNTLEYDLVPLDKNVYVMN
KREELADSKYNERAECEKAVELEQSLDIOTGLDEMAVDQYSLIKDNPRLKRR
HAVLENORTLKAQVLAQADETEGRLNASHVSLBHDYVETGIELTTLVHTMAOEG
VIGARMTGAGGCAIALVOKDVEAFREANKRIEEVYGIASFSYIAEVAAGTIVLD
"
gene
3439..4446
/gene="SP1854"
3439..4446
/gene="SP1854"
/codon_start=1
/translation_table=1
/product="galactose operon repressor"
/protein_id="AAK75926.1"
/db_xref="GI:14973356"
/translation="MATLKDIALQSVSIATVSRVLNRDQSLSTETETRHRLITVAE
LGYTRHLKTGSHKPKOKIAIQWVSQGLDLYYQILGIEKRAOELDDIIRYP
NDHPFTLSEEVIGILICIGKFSRAQISAFEEYQPLVLESDTSLGHTLITDPRAM
KQVVDYPLSGDMRIGILTLEETDQDEILIOKRLFNFKVYSOARGIYHDELIVGR
FTASGYDLMKKEAIOISLDQLPAPFAPSQSLAIGLRALQEGISLPRVSLISND
E"
gene
complement(4492..5529)
/gene="SP1855"
complement(4492..5529)
/gene="SP1855"
/codon_start=1
/translation_table=1
/product="alcohol dehydrogenase, zinc-containing"
/protein_id="AAK75927.1"
/db_xref="GI:14973357"
/translation="MKSAVYTKAGVGLASIRPQIIBADYIIRVVRACYSQDLNR
YRNPETKAGHNSGHEAIGIVEAGEAITYVPGDFVLPFHGGCEDDALLAPDGS
CDNHIGNNIGDFOAEYIRFHYANMALVPIPOPSDTEGMILKSLTTLADYWPCTYHA
ARVANYQKGDYVIGDGAAGVQCAVIAAKMGAQIILMSRHEDRQKAMSGATVAV
AERGQEGITKREILIGGGADALFCVGTAEALIQALGVHNGRKGFPVHYNNRL
GSTFMONISVAGSASATTYDKQFLKAVLQGDINPGHVTSSYKLDIDQAYKDMDE
RKTIKSMIYIE"
gene
complement(5530..5883)
/gene="SP1856"
complement(5530..5883)
/gene="SP1856"
/codon_start=1
/translation_table=1
/product="transcriptional regulator, MexR family"
/protein_id="AAK75928.1"
/db_xref="GI:14973358"
/translation="MNIKSASDLGISADTIRYERYGLVPITRRATGTRDQDDI
EALFEIKCFRSAGVSDVLYMSLYQKDETRERERLGIIEEKOKLEKLSQTLAL
NRLNLIKILYKEGKE"
gene
complement(6127..7026)
/gene="SP1857"
complement(6127..7026)
/gene="SP1857"
/translation="identified by match to Pfam protein family HMM
PF01545"
/codon_start=1
/translation_table=1
/product="cation efflux system protein"
/protein_id="AAK75929.1"
/db_xref="GI:14973359"
/translation="MRNMKAKYAVWVAFLNLTYAIVERIYAGVFGSSAVLADSVHL
GDAIGISAFLETISNRBEDNOYTLGKRKPSLGLAVLYAVLTGSAVLILENTYKI
LHPQVNDIEGILGITITINILASIVYKKTAKNESSLHLEFDLGVAVITLMA
IVRFWDYIIDLPLSLIVSPTLSKALREMTSTKITLDVAPDEIDIKOYSGIERL
DNVASINQNLMTMDLEKNAIYHVCLEMEHMETCKSRIRIFLKDCCGQNTTEIDA
DLETHQTHKRCVDLERSYEHQ"
gene
7150..7686
/gene="SP1858"
7150..7686

```

```

/gene="SP1858"
/codon_start=1
/translation_table=1
/product="transcriptional regulator, TetR family"
/protein_id="AAK75930.1"
/db_xref="GI:14973360"
/translation="MTNDRIRISITKKAIIQAFIQLLMANGYATVQDIIIDADVGR
KNDFYRLKHEHLYSVLADNLKBAHPLPTSYLQHLVMSNFETTLTLMWLKKGQD
FTDDEVQVYIDLLEIPKN"
gene
complement(7853..8662)
/gene="SP1859"
complement(7853..8662)
/gene="SP1859"
/translation="similar to GB:L4203 PID:1161423 PID:1221194
PID:1205327 SP:Q57425; identified by sequence similarity;
putative"
/codon_start=1
/translation_table=1
/product="transporter, putative"
/protein_id="AAK75931.1"
/db_xref="GI:14973361"
/translation="MHHTYQKTIENIKTTLGEMSGYRRVYAAADLFGSTMAIM
DILFAHNSPROMIYLIVLVSFPLMLELVEHRYVDMIGMISLTGIIYIEVSEGRAS
NYLEGLINSYIYLIALQKGYGEVLTTLTYFTVMDPIGLVWVIOAFKKEQEFVAR
KLQKGTWTKYLSISVLMWLAFGFIYOSIGANRPYRDSITDATNGVQGLMTAVYREOW
IFMAATNVFSIYLMGESLQIOGKYLIYILNSLVGMYSKAAKQNTDLN"
gene
complement(9004..10524)
/gene="SP1860"
complement(9004..10524)
/gene="SP1860"
/codon_start=1
/translation_table=1
/product="choline transporter"
/protein_id="AAK75932.1"
/db_xref="GI:14973362"
/translation="MTNLIATPQDSEFPMULRALSCHQLSLTLLAILLAIPLAVEL
RYHEKLDAMVLOIACIPIOTIYSLALGLFPLMGITPLPALTAVIYAIPILQNTY
GIKGDIPNIOEGNIAFGMTRWERKKEFIPLAMPYINSIGRTAVLILIGATLALAG
AGDSFILLGIDRNNAASLILIGLSSAVLIAANFLLKVEKAKRTIFSGRAVAL
LGLISYSPALVQKEKENLVIAKIGIPPELILANMYFLKLEENSTMAATYKPNQCTS
FLYEAIRKKGDDIDYIPEFTGVTESLQSPKVSHPQVYQVARDIAKODHLYLKP"

```

alignment_scores:

Quality:	Ratio:	Length:
Percent Similarity:	51.181	Percent Identity: 28.346

alignment_block:

US-10-048-197-2 x AE007477/rev ..

Align seg 1/1 to reverse of: AE007477 from: 1 to: 12841

```

10 ILSErAlaleuThrAla...LeuLeuValrthrglyCyValserTh 25
   ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6795 CTAGGAGCCTTGCTGTAACAGCTGATCTCGTAACGGGCTGTTCTAGT 6746
   :|||:::GlyAsnValAlaMetLysGluGlnAsnGlnInthrIleGlu 40
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6745 CATTTCGAAATGTCACGACAGATTTTCATCCGACACGATCATGTG 6696
   ::|||:::
40 InThrIlele.....
6695 AGCGGATTCCTGTTAGGAATTATTGGATTACTACATCACTTTGCGC 6646
   :|||:::
44 .....LysGlyLysThrAsnLysGlnLysSerSer 55
6645 AGTCGCTGGTGGTGAAGGAAAGCAAGATGATCTATTCTGAGTCT 6596
   :|||:::
55 gPheGlySerAlaAspSerIleSerPhe.....MetLey 67
   :|||:::
6595 GCATTTCGTGAAGATACGCTAGGCTGAGTATCTATCTATGCGCA 6546

```

```

67 alvalielyphaglyh1sthrAlaileuAlaProAsnArgTrpGln 83
:::|||||
6545 TTGTTCTTCATTTACGACTGCGATATCTCTACATCTCT..... 6508
84 Glu1leuSerleuile1eserPheluTrpValysProGlyArgPr 100
:::|||||
6507 ...CTTTGTCCTTGATCTTCTTTTATCTTTCAAAAGCCCTTCC 6461
100 olysAsnleuSerPhetyleuThralalys 110
:::|||||
6460 ACCT.....TTTGTCTACACTCAAG 6439

seq_name: gb_pr:AC012068

seq_documentation_block:
LOCUS AC012068 166516 bp DNA linear PRI 30-SEP-2000
DEFINITION Homo sapiens BAC clone RP11-196J6 from Y, complete sequence.
ACCESSION AC012068
VERSION AC012068.5 GI:10047984
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE
AUTHORS Andrews, S., Stoneking, T. and Gibson, A.
TITLE The sequence of Homo sapiens BAC clone RP11-196J6
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 166516)
REFERENCE
AUTHORS Waterston, R.H.
TITLE Direct Submision
JOURNAL Submitted (19-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 166516)
REFERENCE
AUTHORS Waterston, R.H.
TITLE Direct Submision
JOURNAL Submitted (26-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 166516)
REFERENCE
AUTHORS Waterston, R.H.
TITLE Direct Submision
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 166516)
REFERENCE
AUTHORS Waterston, R.
TITLE Direct Submision
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 10, 2000 this sequence version replaced gi:9454626.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0196J06

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:

The RPc1-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: PBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-115H13, 200 bp overlap; the clone sequenced to the right is RP11-540C18, 200 bp overlap. Actual start of this clone is at base position 195 of RP11-196J6; actual end is at base position 30203 of RP11-540C18.

FEATURES

source

```

1..166516
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="Y"
/map="Y"
/clone="RP11-196J6"
/clone_id="RPc1-11"
1..328
/note="similar to EST AW964885 (NID:98154616)"
76..336
/note="similar to EST AA325338 (NID:91977613)"
79..468
/note="similar to EST A1471609 (NID:94333699) t166a01.x1"
311..746
/note="similar to EST A1810903 (NID:95397390) tu09a10.x1"
384..727
/note="similar to EST A1134262 (NID:96602449)"
489..746
/note="similar to EST AW367212 (NID:96871862)"
976..1038
/note="similar to EST AW957794 (NID:98147477)"
1061..1108
/rpt_family="L1"
5114..5681
/note="similar to EST AW957794 (NID:98147477)"
5121..5650
/note="similar to EST AA481204 (NID:92210756) aa34d03.r1"
5274..5698
/note="similar to EST AA307574 (NID:91959903)"
5463..5801
/note="similar to EST A1910976 (NID:95630712) wd22a06.x1"
5861..6086
/note="similar to EST BE169475 (NID:98632196)"
6083..6335
/rpt_family="ERV1"
6350..6868
/rpt_family="ERV1"
7040..8710
/rpt_family="ERV1"
8712..8981
/rpt_family="ERVK"
8802..9134
/note="similar to EST A1766349 (NID:95232858) wh06e09.x1"

```



```

misc_feature      1. .87923
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /Chromosome="1"
                    /clone="RP5-100819"
                    /clone_lib="RPC1-5"
                    1. .7745
misc_feature      /note="assembly:fragment:00532
                    fragment.Chain:1"
                    7846. .10386
misc_feature      /note="assembly:fragment:00251
                    fragment.Chain:1"
                    10487. .14930
misc_feature      /note="assembly:fragment:00022
                    fragment.Chain:1"
                    15031. 19573
misc_feature      /note="assembly:fragment:00483
                    fragment.Chain:1"
                    19674. .22034
misc_feature      /note="assembly:fragment:00432
```

misc_feature	fragment_chain:1"	22135_.2496
misc_feature	/note="assembly_fragment:00279	fragment_chain:1"
misc_feature	fragment_chain:1"	24597_.28696
misc_feature	/note="assembly_fragment:00401	fragment_chain:2"
misc_feature	fragment_chain:2"	28797_.37035
misc_feature	/note="assembly_fragment:00446	fragment_chain:2"
misc_feature	fragment_chain:2"	37136_.39184
misc_feature	/note="assembly_fragment:00731	fragment_chain:3"
misc_feature	fragment_chain:3"	39285_.45630
misc_feature	/note="assembly_fragment:00690	fragment_chain:3"
misc_feature	fragment_chain:3"	45731_.47910
misc_feature	/note="assembly_fragment:00012"	48011_.50732
misc_feature	/note="assembly_fragment:00039"	50833_.52856
misc_feature	50833_.52856	/note="assembly_fragment:00087"
misc_feature	52957_.54989	/note="assembly_fragment:00132"
misc_feature	55090_.558682	/note="assembly_fragment:00156"
misc_feature	58783_.63988	/note="assembly_fragment:00206"
misc_feature	64089_.67293	/note="assembly_fragment:00227"
misc_feature	67394_.71607	/note="assembly_fragment:00301"
misc_feature	71108_.74637	/note="assembly_fragment:00472"
misc_feature	74738_.78160	/note="assembly_fragment:00608"
misc_feature	78261_.80922	/note="assembly_fragment:00702"
misc_feature	81023_.87923	

```

BASE COUNT      24752 a 18508 c 17954 g 24603 t 2106 others
ORIGIN
alignment_scores:
    Quality:      76.50      Length:      104
    Ratio:        1.275      Gaps:      3
    Percent Similarity: 57.692      Percent Identity: 25.962
alignment_block:
US-10-048-197-2 x AL356370 ..
Align seg 1/1 to: AL356370 from: 1 to: 87923
19 valthrglycysvalserthrlygsnvalalameltysgulghasncl 35
    ||| ||| ||| ||| |||
55372 GTACCGAGGAGCAAAAGCTGTGGCCCGTGCGGCACCTTGGACCAAAAG 55421
35 nglnthrileguslnthrilelelysglysthrasnlysgln... 50
    ::::::::::::::::::::::::::::||| ||| ||| |||
55422 TGATGTGCTTAGCAAGCAAGCACTAGAGTCTTAACCTATAAACAGTTT 55471
51 .....glulieserSeraArgphceglySerAlaaspserile 62
    ::::::::::::::::::::||| ||| ||| |||
55472 TAGTGCCCTTGGCAATCATATAGCTGCCCAATTGGCCAGT...GATCCCTCTT 55518
63 SerPhemeIleValValIlelySpheglyHis..... 73
    |||::: ::::: ||| ||| |||
55519 TCTTATGTCTTATACCAAGCCCTCTTACACATGACTTGCTTATGCATGC 55568
74 .....ThrAlaIleLeuAlaProbsnArgrtrpglngIleLeuSerL 88

```

[illegible]

FEATURES	*	79101	79100:	gap of	100 bp
source	*	79101	89688:	contig of 10588 bp in length.	
	*	89688	89788:	gap of 100 bp	
	*	89789	103551:	contig of 12563 bp in length.	
			Location/Qualifiers		
			1..102351		
			/organism="Homo sapiens"		
			/db_xref="taxon:9606"		
			/chromosome="1"		
			/map="P32.2-33"		
			/clone="RP5-881J12"		
			/clone_id="RPC1-5"		
			1..10845		
misc_feature			/note="assembly_fragment:00700"		
			clone_end:SP6		
			vector_side:left"		
misc_feature			10946..22122		
			/note="assembly_fragment:00611"		
			fragment_chain:1"		
			22223..35055		
misc_feature			/note="assembly_fragment:00016"		
			fragment_chain:1"		
			35156..50124		
misc_feature			/note="assembly_fragment:00482"		
			fragment_chain:1"		
			50225..60215		
misc_feature			/note="assembly_fragment:00206"		
			fragment_chain:2"		
			60316..66576		
misc_feature			/note="assembly_fragment:00367"		
			fragment_chain:2"		
			66617..70409		
misc_feature			/note="assembly_fragment:00244"		
			70510..79000:		
			/note="assembly_fragment:00869"		
			fragment_chain:3"		
misc_feature			79101..89688		
			/note="assembly_fragment:00779"		
			fragment_chain:3"		
			89789..102351		
misc_feature			/note="assembly_fragment:00186"		
			fragment_chain:3"		
			clone_end:TT7		
			vector_side:right"		
BASE COUNT	29645	a	21820	c	21315
			g	28664	t
					907
					others
ORIGIN					
alignment_scores:					
Quality:	76.50				Length: 104
Ratio:	1.275				Gaps: 3
Percent Similarity:	57.692				Percent Identity: 25.962
alignment_block:					
US-10-048-197-2	x	AL138848	..		
Align seg 1/1	to:	AL138848	from:	1	to: 102351
19	ValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnG1	35			
98031	GTACACGAGGACCAAAAGCTGTGGCCCGCGGCGCCACATTGACCCCAAAATG	98080			
35	nglthrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln....	50			
98081	TGATGTGCTTACCCAGACCTCATTAAGGCTTAACCTATTAACACGTTT	98130			
51GluIleSerArgPheGlySerAlaAspSerIle	62			
98131	TAGTGCTTTGGCAACATATGATGTTCCAGTTTGCCAGT...GATCCTCT	98177			
63	SerPheMetIleValAlaIleLysPheGlyHis.....	73			
98178	TCTATTGCTCTTATCCAAAGCCCTCTAGACACATGACTGCTATGATC	98227			


```

74 .....:ThralleleuAlaProbasnArPrpInguilleUserL 88
          ::|||: ::::: ::::|:||||:||||:|
98228 AAGCCTCACTGACGGCTTACTTCCCATCTCTGGAGTACTGATGACGAAAC 982277
          eu1le1eSerPheutrPvallySPOTyArArProFysanLeuser 104
          || ::||| ||::: ||::: ||:::|:|:|:|:|
98278 TCTACCTGATGCGGGGATTCATCATCGCTTTCCTCCATCATCCCGCAT 98327
          105 pHeTyRLeuThr 108
          ::|||:|:|:|
98328 GATTTCTTGACG 98339
seq_name: gb_pr:AL162430

seq_documentation_block:
LOCUS      AL162430          211791 bp      DNA      linear      PRI 30-NOV-2000
DEFINITION Human DNA sequence from clone RP11-296A18 on chromosome 1, complete
sequence.
ACCESSION  AL162430
VERSION    AL162430
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  Wray,P.
            Direct Submision
            Submitted (30-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequests@sanger.ac.uk
            On Dec 4, 2000 this sequence version replaced gi:11493251.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements. Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information
            on the WormPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/MGP/Chrl
            RP11-296A18 is from the library RPCR-11.2 constructed at the
            Roswell Park Cancer Institute by the group of Pieter de Jong. For
            further details see http://bacpac.med.buffalo.edu/
            VECTOR: PBAC3.6
            This sequence is the entire insert of clone RP11-296A18.

FEATURES
            source
                Location/Qualifiers
                    1..211791
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="1"
                        /clone="RP11-296A18"
                        /clone_lib="RPCR-11.2"
BASE COUNT  61174 a 45199 c 44010 g 61408 t
ORIGIN
Alignment_scores:
Quality:      76.50      Length:      104
Ratio:        1.275      Gaps:      3
Percent Similarity: 57.692      Percent Identity: 25.962

```

[illegible]

Chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>
Rp11-56p10 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone Rp11-56p10 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone Rp11-4H20 is at 125743 in this sequence.
The true right end of clone Rp11-53418 is at 100 in this sequence.

FEATURES

```
source
1. 125842
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="9"
   /map="q31.3-33.3"
   /clone="Rp11-56p10"
   /clone.lib="RPC1-11.1"
   complement(1..98)
   /note="match: GSS: Em:AQ673337"
   102..159
   /note="MIR repeat: matches 1..58 of consensus"
   170..489
   /note="match: GSS: Em:AQ101073"
   1213..1372
   /note="MIR repeat: matches 27..191 of consensus"
   2717..2951
   /note="MIR1A2 repeat: matches 339..566 of consensus"
   3132..3234
   /note="MIR1G repeat: matches 68..166 of consensus"
   3313..3367
   /note="MIR repeat: matches 84..114 of consensus"
   4880..5117
   /note="MIR repeat: matches 2..249 of consensus"
   5505..5659
   /note="L2 repeat: matches 2089..2258 of consensus"
   6076..6140
   /note="MIR repeat: matches 186..247 of consensus"
   6178..6478
   /note="AlusX repeat: matches 1..309 of consensus"
   6575..6861
   /note="AlusC repeat: matches 1..288 of consensus"
   7014..7080
   /note="MIR repeat: matches 18..82 of consensus"
   complement(805..8534)
   /note="match: GSS: Em:B99058"
   8335..8696
   /note="MIR1A2 repeat: matches 1..366 of consensus"
   9093..9338
   /note="HAI repeat: matches 551..806 of consensus"
   9570..9880
   /note="Alu repeat: matches 1..310 of consensus"
   9952..10117
   /note="MIR repeat: matches 10..189 of consensus"
   10433..10794
   /note="match: GSS: Em:AQ076307"
   10523..10712
   /note="L2 repeat: matches 2519..2707 of consensus"
   10847..11025
   /note="MIR repeat: matches 1..179 of consensus"
   11464..11692
   /note="L1 repeat: matches 5318..5560 of consensus"
   12528..12722
   /note="MIR repeat: matches 29..231 of consensus"
   12887..13479
   /note="match: GSS: Em:B71595"
   14082..14207
   /note="MIR repeat: matches 14..140 of consensus"
   14669..14764
   /note="MIR repeat: matches 3..101 of consensus"
   14853..14924
```

```
repeat_region
15221..17233
   /note="L1 repeat: matches 3851..5933 of consensus"
   17392..17465
   /note="L1 repeat: matches 6195..6271 of consensus"
   17726..18368
   /note="L1 repeat: matches 5534..6159 of consensus"
   18371..21223
   /note="L1 repeat: matches 379..3697 of consensus"
   22015..22462
   /note="L2 repeat: matches 2267..2750 of consensus"
   24923..25135
   /note="match: GSS: Em:A260619"
   25144..25424
   /note="match: GSS: Em:A260619"
   25253..25336
   /note="MIR repeat: matches 53..139 of consensus"
   27221..27373
   /note="MIR repeat: matches 98..250 of consensus"
   29278..29540
   /note="AlusX repeat: matches 1..287 of consensus"
   complement(29378..29787)
   /note="match: GSS: Em:AQ375706"
   29541..29572
   /note="8 copies 4 mer gaga 96% conserved"
   complement(29789..30179)
   /note="match: GSS: Em:AQ24025"
   30134..30253
   /note="MIR repeat: matches 56..193 of consensus"
   30414..30521
   /note="MIR repeat: matches 4..114 of consensus"
   30607..30836
   /note="MIR repeat: matches 23..256 of consensus"
   32111..32223
   /note="L2 repeat: matches 2639..2742 of consensus"
   32765..32841
   /note="L2 repeat: matches 2423..2498 of consensus"
   32996..33184
   /note="MIR repeat: matches 1..189 of consensus"
   33275..33575
   /note="MIR repeat: matches 89..386 of consensus"
   33900..33948
   /note="L2 repeat: matches 2702..2750 of consensus"
   34847..35022
   /note="MIR repeat: matches 74..258 of consensus"
   35024..35271
   /note="MIR repeat: matches 17..262 of consensus"
   35487..35799
   /note="AlusX repeat: matches 2..310 of consensus"
   35826..35964
   /note="MIR repeat: matches 98..247 of consensus"
   36356..36430
   /note="MIR repeat: matches 47..130 of consensus"
   36500..36254
   /note="L1 repeat: matches 4386..6143 of consensus"
   38427..38540
   /note="L1 repeat: matches 6042..6157 of consensus"
   38541..38683
   /note="FLAN repeat: matches 1..133 of consensus"
   38684..38699
   /note="L1 repeat: matches 6027..6042 of consensus"
   38700..38994
   /note="Alu repeat: matches 4..298 of consensus"
   38995..39666
   /note="L1 repeat: matches 5369..6027 of consensus"
   39668..39712
   /note="L1 repeat: matches 5322..5365 of consensus"
   39709..41178
   /note="L1 repeat: matches 4204..5645 of consensus"
   41179..42394
   /note="L1 repeat: matches 4121..5332 of consensus"
   42395..42697
   /note="MIR repeat: matches 1..320 of consensus"
```


[illegible]

		<p> GPTKILMHLPDPEVEYRILVTLRPGEGGTGPPGGLPTLRKCADPVGPNVLEVDLR AROTLWMLPEPAYSVTRCHSNLTLSEGTGSPGKREYAEAEVLTSHYTLGLRPT AVYLRILACNPEGRKESEGETIYVHDEEDVPPILETSGGPFEEKIYLOMKAPENNG VISMETYSYKAVGSLDPSADLSNOKGIVLRLRNTHLPGLLPGSTSTTRASTKSG GFSPEYVARIATKIIAASPLTDSAEAGEGLNEEPTIYMLPAPSRGAPVAYOLVKE ERAORSGRADGGGCFPPVPGYRNASMLSESHVFAELKPLQALVTPPEFVGNRTYN GYAMPPLSPVASYSTIYFQALSTANGETIKESCHVLASGASTONSNTVEPKPADSTVO TAGVIAGMLMEIILLGAMMTIKRRNYSYSYKLAKKNGKOSGOREMGLVASDK SSSLSSTRNEDPPEFSCODVNGFSTSPCSFSYOKLKKOKLSLSCYSNSHYDMNQ PTLTIQHTPRTECPLEMAYPROGLQPAIRVADLQIITOMKRGQCFREVEALPE GQTSWPTAKEDENKRNRRYGNITSYSHSRRLQOILEDSHVRSTINANYVDGHRPY YIAQSGMOETKDFKPMRIMQENSI SVYAMTNLAEVGRVADCPADPDTEIVGDIKYS ALENEPLAEYIIRFTVQKSGHEIRIEICQVFLASGDSNDNPGVATGLGFIQVDFL NPDPAGITVHSCAGRTGCFEIIDMLMAEHEGVDJFNCVRELRAORVMVQRE EQYFVDAITLEACLTGNTAIPVCEFSVYIYINISBMDPQINCSQINDEP" 1396. >4920 </p>							
mat_peptide		/gene="ptprt"							
		/product="receptor protein tyrosine phosphatase rho"							
misc_feature	1399..1887	/gene="ptprt"							
		/note="Region: MAM domain"							
misc_feature	1933..2115	/gene="ptprt"							
		/note="Region: Ig-like domain"							
misc_feature	2125..2178	/gene="ptprt"							
		/note="Region: fibronectin type III repeat"							
misc_feature	2464..2730	/gene="ptprt"							
		/note="Region: fibronectin type III repeat"							
misc_feature	2767..3042	/gene="ptprt"							
		/note="Region: fibronectin type III repeat"							
misc_feature	3088..3336	/gene="ptprt"							
		/note="Region: fibronectin type III repeat"							
misc_feature	3217..3228	/gene="ptprt"							
		/note="proteolytic subtilisin/furin cleavage site"							
misc_feature	3559..3621	/gene="ptprt"							
		/note="Region: transmembrane domain"							
misc_feature	4504..4809	/gene="ptprt"							
		/note="probably catalytic; Region: phosphatase domain"							
BASE COUNT	1290 a 1276 c 1190 g 1151 t	13	others						
ORIGIN									
<p> alignment_scores: <table border="0" style="width: 100%;"> <tr> <td>Quality: 75.00</td> <td>Length: 90</td> </tr> <tr> <td>Ratio: 1.562</td> <td>Gaps: 4</td> </tr> <tr> <td>Percent Similarity: 53.333</td> <td>Percent Identity: 30.000</td> </tr> </table> </p>				Quality: 75.00	Length: 90	Ratio: 1.562	Gaps: 4	Percent Similarity: 53.333	Percent Identity: 30.000
Quality: 75.00	Length: 90								
Ratio: 1.562	Gaps: 4								
Percent Similarity: 53.333	Percent Identity: 30.000								
<p> alignment_block: <table border="0" style="width: 100%;"> <tr> <td>US-10-048-197-2</td> <td>x</td> <td>AFI173857</td> <td>..</td> </tr> </table> </p>				US-10-048-197-2	x	AFI173857	..		
US-10-048-197-2	x	AFI173857	..						
<p> Align seg 1/1 to: AFI173857 from: 1 to: 4920 </p>									
15	ThrlAlaLeuIeuValThrGlyCysValSerThrGlyAsnValAlaMetIy 31								
2017	ACGGCCCTCATGTCACACGTCGCGG..... 2043								
31	seLugInaengInGlnThrIleGluInThrIleIleIySgIyLysTha 48								
2044AACCAAAACGCTCTCACGCCACCGTCAGTGTGGGAGACAT 2086								
48	snLysGlnGluIleSer.....SerArgPheGly 57								
2087	CITCAGCGCATCATACCAATACCGTGTCCTTCGCTTCACAGGGGA 2136								
58	SerAlaAspSerIleSerPheMetIleValIleIleYsPheGlyHisTh 74								

2137 TCAGAGTCTCC...AACTATGACAGACCACTGTCGTAAGAGCCCTCCAC 2183

74 ralaileuallaproanargtrpglnlileuserleuallies 91
 |||
 |||
 2184 CCAATTCACCCCT.....GAGCTCTGCTGTGGGGCCA 2221
 |||
 |||
 91 erpheleutrpvallyspro 97
 :|||
 :|||
 2222 CCTACTGTGATCAGCCA 2241

seq_name: gb_pr:AL512598

seq_documentation_block:
 LOCUS AL512598 139491 bp DNA linear PRI 13-NOV-2001
 DEFINITION Human DNA sequence from clone RP11-165A20 on chromosome 10,
 complete sequence.
 ACCESSION AL512598 AC024986
 VERSION AL512598.11 GI:16944073
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 139491)
 Ashwell, R.
 DIRECT SUBMISSION
 Submitted (09-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk
 On Nov 15, 2001 this sequence version replaced gi:16151438.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human
 chromosome 10, constructed by the Sanger Centre Chromosome 10
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr10
 RP11-165A20 is from the library RPCI-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.choi.org/bacpac/home.htm
 VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone
 RP11-165A20. It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true left end of clone RP11-165A20 is at 1 in this sequence.
 The true left end of clone RP11-80K21 is at 137492 in this
 sequence. The true right end of clone RP11-12907 is at 37219 in
 this sequence.

FEATURES
 source
 Location/Qualifiers
 1..139491
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-165A20"
 /clone_lib="RPCI-11.1"
 BASE COUNT 40284 a 28155 c 27520 g 43532 t
 ORIGIN

alignment_scores:
 Quality: 74.50 Length: 69
 Ratio: 1.910 Gaps: 3
 Percent Similarity: 56.522 Percent Identity: 30.435

alignment_block:
 US-10-048-197-2 x AL512598 ..
 Align seg 1/1 to: AL512598 from: 1 to: 139491

43 lileysglysThrAsnLysGlnLileSerSerarPheglySerAl 59
 |||
 |||
 19557 ATATCCAGCGAAGACAGAGACAG.....ACCTG 19585
 59 aaspSerLileSerPheMetLileValLileLysPheGlyHisThrAlar 76
 |||
 |||
 19586 CGATGTGCTA.....TTAAAGTCATACAAAGTATGCGCATCTGGGC 19629
 76 lileuAlapro.....AsnArgTrpGlnLileLileSerLeu 88
 :|||
 :|||
 19630 TCATGTTCTTTCAAGAACTCTCCAGTGGGATATATGACGCTTC 19679
 89 lileSerPheLeuTrpVallysProTyraRgProLysAsnLeuSerph 105
 :|||
 :|||
 19680 ACATTAGAGTCTATGCTTGACCTTGAGGAACCTGGTTCCTGTGT 19729
 105 eTyrlLeu 107
 |||
 |||
 19730 TTATTTG 19736

seq_name: gb_htg:AP001795

seq_documentation_block:
 LOCUS AP001795 143873 bp DNA linear HTG 30-MAY-2000
 DEFINITION Homo sapiens chromosome 11 clone RP11-711B7 map 11q23. WORKING
 DRAFT SEQUENCE, 19 unordered pieces.
 ACCESSION AP001795
 VERSION AP001795.2 GI:8117467
 KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
 SOURCE Homo sapiens DNA, clone:RP11-711B7.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 143873)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,T.
 Homo sapiens 143,873 genomic DNA of 11q23
 Published Only in database (2000) in press
 2 (bases 1 to 143873)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,T.
 Direct Submission
 Submitted (11-APR-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
 Japan (E-mail:hattori@gsc.riken.go.jp,
 URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
 Fax:81-42-778-9924)
 On May 30, 2000 this sequence version replaced gi:7592910.

COMMENT
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center(GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: RP11-711B7
 Center clone name: RP11-711B7
 ----- Summary Statistics
 Sequencing vector: PCR products, 100% of reads
 Chemistry: Dye-terminator ET-amersham, 100% of reads
 Assembly program: Phrap, version 0.990329
 Consensus quality: 134262 bases at least Q40


```

46615 .....CTGTCAATCCATCATACACCAAGATT 46587
71 egiyhsthrAlaileuAlaProasnaGrpInguilileuSerI 88
46586 TGGCCATACCTGTCTCATTT.....ACCTTTGTTGTTGTACATACAC 46543
88 eulileleserPhelEtrPVallysProTyraRgProLysaLseuSer 104
46542 ATGTTATGATTTCATTTGGCGTTTCCTTATCCCTTTCCAGCTTACGA 46493
105 Phe 105
46492 TTT 46490

seq_name: gb_pl:AP003215

seq_documentation_block:
LOCUS AP003215 154137 bp DNA linear PLN 04-AUG-2001
DEFINITION Oryza sativa genomic DNA, chromosome 1, BAC clone:OSJNBa0089K24.
ACCESSION AP003215
VERSION AP003215.3 GI:15128217
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (cultivar:Nipponbare) DNA, clone:OSJNBa0089K24.
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehmerioideae; Oryzaceae; Oryza.
1 (bases 1 to 154137)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
clone:OSJNBa0089K24
2 (bases 1 to 154137)
Published Only In Database (2001) In press
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rsgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Aug 9, 2001 this sequence version replaced gi:13442957.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI NonRedundant Protein database, nr
(ffp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologies of the coding regions were searched against
NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding DBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted with a gene prediction program
is classified as a 'hypothetical' protein.
The orientation of the sequence is from -21M13 to M13Rev of the BAC
clone. This sequence of OSJNBa0089K24 clone has an overlap with
P0013F10 clone (DBJ: AP002523) at the position 151,639 to 154,137
of 3' end. The sequence of this clone ends at the position 2,499 of
P0013F10. Detailed information on overlap and assembly quality
together with annotation of this entry is available at
http://rsgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES
Source
1. 154137
/oranism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="OSJNBa0089K24"
join(4692..4907,4987..5166,5273..5414,6051..6472)
/gene="OSJNBa0089K24.1"
gene

```

```

CDS
join(4692..4907,4987..5166,5273..5414,6051..6472)
/gene="OSJNBa0089K24.1"
/codon_start=1
/product="putative peroxidase"
/protein_id="BAB62546.1"
/db_xref="GI:15128218"
/translation="MAISEMSALFELFSALRSLSVHSGQLGIFYDNNCPDAEIVR
STVEKYYNNATITAPGLRLHFHDFCPGGDASVLSGASERTAPQNFIRGEVID
DAKSQILAVCSGVSCADILALARDADVLTGGSNVPGLRRGRSSASDAKALPS
PADPVSVQRQKFAAOGJLDRNHTIGDTDCIFFRYLRLNATGADPTISPSALPOLR
ALCPADGDSRRVALDGSFGAFDVSEFKVNRDGAVALSDQRLMGAAQAQAAVSFA
GNVRLGFLRSEYFEPRKAWMSIAVTSQGEIRRKCKFN"
8547..9122
/gene="OSJNBa0089K24.2"
8547..9122
/gene="OSJNBa0089K24.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB62547.1"
/db_xref="GI:15128219"
/translation="MARRRPPQRHHHLVLLALAFSCLASGRTATPTPTSSOQPO
PAGTLQPALADLAARTCMVTVOIKTSCASPMRTSDAVSLAREDAYRLNGLARLG
SSSPDPSAFAERCAATDIFRVGPGGVCYLIRRSRGDWTQMVRYEPTSDT
PSFRYGDPLPNVAVGFNRCPRLAAGAAAO"
complement(join(13385..13475,13662..13754,13888..13995,
14101..14158,14320..14554))
/gene="OSJNBa0089K24.3"
complement(join(13385..13475,13662..13754,13888..13995,
14101..14158,14320..14554))
/note="hypothetical protein
similar to Arabidopsis thaliana chromosome 1, F6D8.19"
/codon_start=1
/protein_id="BAB62548.1"
/db_xref="GI:15128220"
/translation="MATTSAACYPRLSAPPULHRRKGPPOOQLRTAVGPRRRPSIQ
VAAAARAGQSVAAATDAEFQSDRPMLFPGVNCILNGVRFYOEADPNSIY
IPLOSSEGRKLORSGRAPDISVSEVDEKRSYKEAVALRMINGVILNLPQLATPL
NVPVLFDRADVADNARYAVGRSSECELI"
complement(join(15491..15586,15712..15776,15919..15981,
16399..16647,16744..16929))
/gene="OSJNBa0089K24.4"
complement(join(15491..15586,15712..15776,15919..15981,
16399..16647,16744..16929))
/note="contains ESTs ANU00786(C50379),C26898(C50379)
similar to Arabidopsis thaliana chromosome 1, F28N24.7
unknown protein"
/codon_start=1
/protein_id="BAB62549.1"
/db_xref="GI:15128221"
/translation="MAYETTEGVRNLAVEGPEPAAAAAAGGGGEGNORRASSNR
IQVSNTRKPLFFVYNLAKRTYMOCHGDELVSALGMATVYTVVAEILKNNGEPAEKIR
TSVVEINDESRRVPLQAKTEIVLESKEKDELMAAAREEAEEEOA"
complement(join(21465..22160,21730..21828,21923..21995,
22080..22190,22287..22460,22570..23666,23741..25162))
/gene="OSJNBa0089K24.5"
complement(join(21465..21650,21730..21828,21923..21995,
22080..22190,22287..22460,22570..23666,23741..25162))
/note="hypothetical protein
similar to Oryza sativa chromosome 1, P0489A01.2"
/codon_start=1
/protein_id="BAB62550.1"
/db_xref="GI:15128222"
/translation="MADRDEQILYDTIAGSSQYVNEEGNEDPNQYLNESGNERD
AENQOGHVERDVEGNOEERASGSPSVGQRKARGAKKLEGRHITTEVEDERP
SAPAEAKNRYRSGWNRDYNVSVYRHRTRARGDHESEKEMLWTMLETF
TLPAGETDKVKTWTLKMAKPOFOSGDKTLKYLKGQNFPTPKLRDHPAEFAY
KTGEOGKAMERKNNAKAKKHNNHSGSGSYVAMPKWEIIEASILERGIEPTAMP
ERSKFTYIAHGTLNADSGSLVGYOQEAARKJLTVVEASSGCTRRPRDRBELTIA
LQTPHGRGRTKRGVLPWKIGFKEDITHYTSRRRSKRDTEAKTADLEFRVSTELNMQ
EVARKVDERMAHRSHDPTLIPVWVSFSGNRRSCASGVGSGSMDAMOTODEST

```


TITLE
JOURNAL
COMMENT

Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.-D., Young, G.,
Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (16-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIMR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L17460
Center clone name: 350_G_1
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161250 bases at least Q40
Consensus quality: 16516 bases at least Q30
Consensus quality: 167502 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 168260; sum-of-ctrls
Quality coverage: 12.7 in Q20 bases; agarose-fp
Quality coverage: 14.2 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 78525: contig of 78525 bp in length
* 78526 78625: gap of 100 bp
* 78626 79185: contig of 560 bp in length
* 79186 79285: gap of 100 bp
* 79286 79495: contig of 210 bp in length
* 79496 79595: gap of 100 bp
* 79596 80537: contig of 942 bp in length
* 80538 80637: gap of 100 bp
* 80638 81632: contig of 995 bp in length
* 81633 81732: gap of 100 bp
* 81733 82472: contig of 740 bp in length
* 82473 82572: gap of 100 bp
* 82573 83248: contig of 676 bp in length
* 83249 83348: gap of 100 bp
* 83349 84207: contig of 859 bp in length
* 84208 84307: gap of 100 bp
* 84308 84514: contig of 207 bp in length
* 84515 84614: gap of 100 bp
* 84615 85295: contig of 681 bp in length
* 85296 85395: gap of 100 bp
* 85396 86100: contig of 705 bp in length
* 86101 86200: gap of 100 bp
* 86201 86869: contig of 669 bp in length
* 86870 86969: gap of 100 bp
* 86970 87738: contig of 769 bp in length
* 87739 87838: gap of 100 bp
* 87839 88571: contig of 733 bp in length
* 88572 88671: gap of 100 bp
* 88672 89751: contig of 1080 bp in length
* 89752 89851: gap of 100 bp
* 89852 90536: contig of 685 bp in length
* 90537 90636: gap of 100 bp
* 90637 91599: contig of 963 bp in length
* 91600 91699: gap of 100 bp
* 91700 92353: contig of 654 bp in length

* 92354 92453: gap of 100 bp
* 92454 93487: contig of 1034 bp in length
* 93488 93587: gap of 100 bp
* 93588 94231: contig of 644 bp in length
* 94232 94331: gap of 100 bp
* 94332 95444: contig of 1113 bp in length
* 95445 95544: gap of 100 bp
* 95545 96659: contig of 1115 bp in length
* 96660 96759: gap of 100 bp
* 96760 97953: contig of 1194 bp in length
* 97954 98053: gap of 100 bp
* 98054 98770: contig of 717 bp in length
* 98771 98870: gap of 100 bp
* 98871 99995: contig of 1125 bp in length
* 99996 100095: gap of 100 bp
* 100096 100895: contig of 800 bp in length
* 100896 100995: gap of 100 bp
* 100996 101656: contig of 661 bp in length
* 101657 101756: gap of 100 bp
* 101757 102779: contig of 1023 bp in length
* 102780 102879: gap of 100 bp
* 102880 104345: contig of 1466 bp in length
* 104346 104445: gap of 100 bp
* 104446 105396: contig of 951 bp in length
* 105397 105496: gap of 100 bp
* 105497 106713: contig of 1217 bp in length
* 106714 106813: gap of 100 bp
* 106814 108439: contig of 1626 bp in length
* 108440 108539: gap of 100 bp
* 108540 109727: contig of 1188 bp in length
* 109728 109827: gap of 100 bp
* 109828 148957: contig of 39130 bp in length
* 148958 149057: gap of 100 bp
* 149058 171660: contig of 22603 bp in length.

FEATURES

source

1. 171660
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_id="RP24-350G1"
/clone_id="RPC1-24 Male Mouse BAC"
1. 78525
/note="assembly-fragment
clone_end:sp6
vector_side:left"
78626. 79185
/note="assembly-fragment"
79286. 79495
/note="assembly-fragment"
79596. 80537
/note="assembly-fragment"
80638. 81632
/note="assembly-fragment"
81733. 82472
/note="assembly-fragment"
82573. 83248
/note="assembly-fragment"
83349. 84207
/note="assembly-fragment"
84308. 84514
/note="assembly-fragment"
84615. 85295
/note="assembly-fragment"
85396. 86100
/note="assembly-fragment"
86201. 86869
/note="assembly-fragment"
86970. 87738
/note="assembly-fragment"
87839. 88571
/note="assembly-fragment"
88672. 89751
/note="assembly-fragment"
89852. 90536
misc_feature


```

* 42753 68827: contig of 26075 bp in length
* 68828 68927: gap of 100 bp
* 68928 98647: contig of 29720 bp in length
* 98648 98747: gap of 100 bp
* 98748 126143: contig of 27396 bp in length
* 126144 126243: gap of 100 bp
* 126244 172182: contig of 45939 bp in length.
Location/Qualifiers
1. 172182
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="RP11-4E13"
/clone_id="RP11-11 Human Male BAC"
1. 5011

```

```

misc_feature
/note="assembly_fragment"
5112..9430
/note="assembly_fragment"
9531..14192
/note="assembly_fragment"
14293..20845
/note="assembly_fragment"
20946..30131
/note="assembly_fragment"
clone_end:SP6
vector_side:right
30232..42652
/note="assembly_fragment"
42753..68827
/note="assembly_fragment"
68928..98647
/note="assembly_fragment"
clone_end:r7
vector_side:right
98748..126143
/note="assembly_fragment"
126244..172182
/note="assembly_fragment"
BASE COUNT 53360 a 30939 c 31936 g 55042 t 905 others
ORIGIN

```

```

alignment_scores:
Quality: 74.50 Length: 85
Ratio: 1.490 Gaps: 3
Percent Similarity: 58.824 Percent Identity: 25.882

```

```

alignment_block:
US-10-048-197-2 x AC021986/rev ..

```

```

Align seg 1/1 to reverse of: AC021986 from: 1 to: 172182

```

```

22 CysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrI 38
|||||
134156 TGTGCTGCCACAGGGTTTGTGTGTCAGACAGCATTCAGCCAGCAT 134107
|||||
38 eglu glnThrIleIleLysGlyLysThrAsnLysGlnGlnIleSerSer 54
|||||
134106 TGAGGGCTTCATCATCACAGTCTTCATTCGACAGTTATCTT..... 134063
55 ArgpNegLysSerAlaAspSerIleSerPheMetIleValIleLysPh 71
|||||
134062 .....CTGTCTATTCCTCATCATCACCAACAGATT 134034
71 eglYHISThrAlaIleLeuAlaProAsnArgTrpGlnGlnIleLeuSerL 88
|||||
134033 TGGCCATACCTGTGTGTCATT.....ACCTTGTGTGTGTGTGCATACAC 133990
88 euIleIleSerPheLeuTrpValLysProTyrArgProLysAsnLeuSer 104
|||||
133989 AGCTATGATATTACATTGGGCTTTCTTATTCCTTTCCACGTTAGCA 133940
105 phe 105
|||

```

```

133939 TTT 133937
seq_name: gb_PR:AC019184

```

```

seq_documentation_block: 194372 bp DNA linear PRI 09-MAY-2001
LOCUS AC019184
DEFINITION Homo sapiens BAC clone RP11-298E6 from 2, complete sequence.
ACCESSION AC019184
VERSION AC019184.3 GI:14010913
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 194372)
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
2 (bases 1 to 194372)
Meyer, R., Maupin, R. and Bolla, S.
The sequence of Homo sapiens BAC clone RP11-298E6
Unpublished
3 (bases 1 to 194372)
Waterston, R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
MO 63108, USA
4 (bases 1 to 194372)
Waterston, R.
Direct Submission
Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 9, 2001 this sequence version replaced gi:17109570.
----- Genome Center

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0298E06

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Firengen, E., Tateo, M., Catanesi, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://dnapac.med.buffalo.edu>)
VECTOR: pBac3.6

38 eglu.GlntHrlellelsglyLysGlyStHAsysIngluileSerSer 54
||||| ||||| : : : : :
55341 TGAGGCCCTTTCATCATCACAGCTCTTCTCATCGAAATTATCTT 55384

55 ArgpHeGlySerAlaAspSerIleSerPheMetIleValIleLysPhe 71
: : : : :
||||| : : : : :
55414 TGCCCATACCTGGTCAT ACCCTTGCTGTTGTTGTAACATACAC 55457

88 euIllelSerPheLeuValLysProTyrArgProLysAsnLysSer 104
: : : : :
55458 ATGTTATGATTTCATATTGGCTTTCTTCCTTATTCCTTTCCACGTTAGCA 55507

105 phe 105
|||
55508 TTT 55510

seq_name: gb_pr:AC062004

seq_documentation_block:
LOCUS AC062004 205646 bp DNA linear PRI 06-OCT-2001
DEFINITION Homo sapiens chromosome 8, clone RP11-325020, complete sequence.
ACCESSION AC062004
VERSION AC062004.5 GI:15982523
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 205646)
Biren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-325020
Unpublished
2 (bases 1 to 205646)
Biren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedal,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,T.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGurt,A., McKernan,K., McPheters,R.,
Meldrum,J., Menes,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., P.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Kogov,P., Rothman,D.,
Roy,A., Santos,N., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,R., Stojanovic,N., Subramanian,A., Talamas,J.,
Teefaye,S., Theodore,J., Tirrell,A., Travers,W., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (21-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 205646)

TITLE
JOURNAL
REFERENCE
AUTHORS

3
Biren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J.Y., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Goid,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Huine,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,

```

TITLE
JOURNAL
COMMENT
FEATURES
source
-----
Location/Qualifiers
1..205646
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-325020"
/clone.lib="RPCT-11 Human Male BAC"
583..613
/rpt_family="AT_rich"
1040..1082
/rpt_family="(7)n"
complement(2328..2465)
/rpt_family="HALb"
complement(2521..2963)
/rpt_family="MT1c"
3087..3097
/rpt_family="AT_rich"
complement(3098..3387)
/rpt_family="AluX"
3388..3410
/rpt_family="AT_rich"
3524..3573
/rpt_family="AT_rich"
4752..4793
/rpt_family="AT_rich"
/rpt_family="AT_rich"
complement(5166..5170)
/note="<30 qual SNGL region"
5528..5951
/rpt_family="MT1A1"
complement(6207..6289)
/rpt_family="L1M4"
6292..6325
/rpt_family="MT1A1"
6326..6360
/rpt_family="L1M4"
6361..6846
/rpt_family="CA)n"
/rpt_family="MT1A1"
complement(6996..7320)
/rpt_family="MT1A1"
complement(7333..8629)
/rpt_family="L1M4"
9323..9384
/rpt_family="(TA)n"
complement(9459..9638)

```

```

repeat_region      /rpt_family="L1M4"
complement(9806..10824)
repeat_region      /rpt_family="L1Mda"
10826..11001
repeat_region      /rpt_family="(TA)n"
11009..11167
repeat_region      /rpt_family="(CATATA)n"
complement(11176..11527)
repeat_region      /rpt_family="L1Mda"
complement(11728..12018)
repeat_region      /rpt_family="L1MEC"
12123..12159
repeat_region      /rpt_family="AT_rich"
12423..12734
repeat_region      /rpt_family="Alusg"
12968..13048
repeat_region      /rpt_family="L2"
14102..14245
repeat_region      /rpt_family="MIR"
14332..14359
repeat_region      /rpt_family="AT_rich"
14370..14543
repeat_region      /rpt_family="L1ME3A"
14562..14734
repeat_region      /rpt_family="L1ME3A"
14749..14788
repeat_region      /rpt_family="AT_rich"
complement(14992..15305)
repeat_region      /rpt_family="AluY"
complement(15585..15927)
repeat_region      /rpt_family="MER7A"
15929..15965
repeat_region      /rpt_family="(TG)n"
16097..16124
repeat_region      /rpt_family="AT_rich"
16242..16262
repeat_region      /rpt_family="AT_rich"
16272..16726
repeat_region      /rpt_family="L1PA5"
16727..17021
repeat_region      /rpt_family="L1PA4"
17022..17110
repeat_region      /rpt_family="L1PA5"
22465..22497
repeat_region      /rpt_family="AT_rich"
22801..23094
repeat_region      /rpt_family="MLT1J2"
23609..23630
repeat_region      /rpt_family="AT_rich"
23688..24063
repeat_region      /rpt_family="MSTA"
24066..25688
repeat_region      /rpt_family="MSTA-int"
25689..25753
repeat_region      /rpt_family="MSTA"
complement(25751..26310)
repeat_region      /rpt_family="rigger1"
26311..26801
repeat_region      /rpt_family="MER9"
complement(26802..26822)
repeat_region      /rpt_family="rigger1"
complement(26952..27270)
repeat_region      /rpt_family="rigger1"
27721..27999
repeat_region      /rpt_family="AlusC"
complement(28000..28181)
repeat_region      /rpt_family="rigger1"
complement(28322..29029)
repeat_region      /rpt_family="L1MB8"
29037..29105
repeat_region      /rpt_family="(TA)n"
29865..29894
repeat_region      /rpt_family="AT_rich"

```

```

repeat_region      30308..30347
repeat_region      /rpt_family="(TA)n"
complement(31612..31739)
repeat_region      /rpt_family="MIR3"
32004..32169

alignment_scores:
    Quality: 74.50      Length: 97
    Ratio: 1.380      Gaps: 5
    Percent Similarity: 35.670      Percent Identity: 28.866

alignment_block:
US-10-048-197-2 x AC062004/rev ...

Align seg 1/1 to reverse of: AC062004 from: 1 to: 205646

22 CysValSerThrGlyAsnValAlaMetLysGluInaGln.....36
||| ||||| ||| :||| ||||| |||
46222 TGTGTTCCACAGAAAC.....TTAAGGTCCTAACTCATATGCGC 46179

37 .....ThrIleGluInThrIleIleLysGlyLysThra 48
|||:||||| ||| ||||| ||| ||| |||
46178 AGTTCCTTCATTCCTCATTCGACGACCTCATTAATTCAGATCA 46129

48 subysGlnGluIleSerSerArgPheGlySerAlaAspSerIleSerPhe 64
:||||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
46128 AAAGAAAGAA.....AGCAGTCTACTCTTTT 46097

65 MetIleValValIleLysPheGlyHisThrAlaIleLeu..AlaProAs 80
|||||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
46096 TTTCCTCTTTTAAAGATGAGGCTCGCTGTGCGCAGCTGAGT 46047

80 naGTTPGInGluIleLeuSerIleIleSerPheLeuTrpValLysP 97
|||||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
46046 GCAGTGGCAGGATTATCTCGACTGCTTACTATA.....ACCTCA 46006

97 rotYrArpProLysAsnLeuSerPheTyrLeuThrAlaLys 110
|||||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
46005 ACACACAGCCTAAAGCATCTGTACTTGTGCTGCCAA 45965

seq_name: gb_ba:AF253043

seq_documentation_block:
LOCUS AF253043 1377 bp DNA linear BCF 03-NOV-2000
DEFINITION Lactobacillus crispatus strain LMG 12003 surface layer protein
(slpA) gene, complete cds.
ACCESSION AF253043
VERSION AF253043.1 GI:7739699
KEYWORDS
SOURCE
ORGANISM Lactobacillus crispatus.
Lactobacillus crispatus.
Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
Lactobacillus.
REFERENCE 1 (bases 1 to 1377)
AUTHORS Sillanpaa,J., Martinez,B., Antikainen,J., Toba,T., Kalkkinen,N.,
Tankka,S., Lounasma,K., Keranen,J., Hook,M.,
Westerlund-Wikstrom,B., Pouwels,P.H. and Korhonen,T.K.
Characterization of the collagen-binding S-layer protein CbsA of
Lactobacillus crispatus
J. Bacteriol. 182 (22), 6440-6450 (2000)
11053389
JOURNAL MEDLINE
20507814
PUBMED 2 (bases 1 to 1377)
AUTHORS Martinez,B., Smit,E. and Pouwels,P.H.
DIRECT SUBMISSION
JOURNAL Submitted (06-APR-2000) IPLA-CSIC, Carretera de Infesto s/n,
Villaviciosa 33300, Spain
FEATURES
SOURCE
1..1377
/organism="Lactobacillus crispatus"
/strain="LMG 12003"
/db_xref="taxon:47770"
1..1377
gene

```

```
CDS
    /gene="sipNA"
    1..1377
    /gene="sipNA"
    /note="SipNA"
    /codon_start=1
    /product="surface layer protein"
    /protein_id="AAF68971.1"
    /db_xref="GI:7739700"
    /translation="MKKLRIVYSAAAALLAVPVAASVSNADSTATTATTANADY
    /TINIGGTIVSNENOVDPSPISVLGNVKNKTAGNLISKATLSGTSATFGKSTAN
    LFGTDQNNVLINGRPAKDELISVNASSTNITIKNGVFNPEKNGKTTLVSSNKV
    TFGSDNAKTVISLDONGTAKDLTVNISDVAFNATNGVVEVNTTGTQAHAGNAM
    VLAANGOLNTAALPAIESNVAVORVDSNGNGTVNADFKEHVNNIEPATAKD
    QIAKONIDYGPQGFEPKAPRPTFYKATSSINGKSELPETPEVTAVNADPVPSQPKT
    IAHNATYKEDGTTNRANDKARKYESVNVAMSKTKIGKNDPFEVITDKGATGATYIAD
    NIDGTRKRLKHAHYVYKTSKRANKVVLKRGDVIYVGGTYFFKNGKYKITYNTEK
    TVYKASNF"

BASE COUNT      445 a      276 c      261 g      395 t
ORIGIN

alignment_scores:
    Quality:      74.00      Length:      81
    Ratio:        1.451      Gaps:        3
    Percent Similarity: 62.963      Percent Identity: 27.160

alignment_block:
US-10-048-197-2 x AF253043 ..
Align seg 1/1 to: AF253043 from: 1 to: 1377

1 MetLeuAaNHSHisHisIleArgLeuThrIleSerAlaLeuThrAlaIle 17
  :::|||||.....:|||||:::|||||.....:
202 GTCTTAATGCTAATGTCAGAAATACGCTGTAACCTTATTTCAAAGCC 251
  17 uleuValThrGlyCysValSer.....Thrg 26
  :::::|||||:::|||||
252 TACTTTGAGTGCTGTTCTATTTCAGCAACTTCGGTGCAGAACGACCTACG 301
  26 IyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluGlnThrIle 42
  :::::|||||:::|||||
302 CTACTTAAAGAGTACTGACCAAAAC.....AACGTTTAA 336
  43 IleLysGlyLysThrAsnLysGlnGluIleSerSer...ArgPheGlyIse 58
  ||:::|||||:::|||||
337 ATTAAAGGTAGAACGCTAAAGATCAATTAGTAACTGTAATGCTGCGCAG 386
  58 fAlaAspSerIleSerPheMetIleValIleLysPheGly 72
  ||:::|||||:::|||||
387 CTCACACACTATTACTCAAAATGTTGATTTAACTTTGCG 429
  seq_name: gb_ba:AF421351

seq_documentation_block:
LOCUS      AF421351      2884 bp      DNA      linear      BCT 18-OCT-2001
DEFINITION Azotobacter vinelandii L-isoaspartate o-methyltransferase (pcm)
            gene, partial cds; and murein endopeptidase (mep) and stationary
            phase sigma factor (rpos) genes, complete cds.
ACCESSION  AF421351
VERSION    AF421351.1  GI:16226162
SOURCE     Azotobacter vinelandii.
KEYWORDS   Azotobacter vinelandii.
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Azotobacter.
REFERENCE  1 (bases 1 to 2884)
            Kujat Choy,S.L., Weakins,D., Tindale,A.E. and Page,W.J.
            The stationary phase sigma factor (rpos) of Azotobacter vinelandii
            JOURNAL Unpublished
            2 (bases 1 to 2884)
            Tindale,A.E., Weakins,D., Kujat Choy,S.L. and Page,W.J.
            Direct Submission
```

```
JOURNAL      Submitted (19-SEP-2001) Biological Sciences, University of Alberta,
              CW405 Biological Sciences Bldg., Edmonton, Alta T6G 2E9, Canada
FEATURES
source
    organism="Azotobacter vinelandii"
    strain="UM"
    /db_xref="taxon:354"
    <1..396
    /gene="pcm"
    <1..396
    /gene="mep"
    /gene="rpos"
    /codon_start=1
    /transl_table=11
    /product="L-isoaspartate o-methyltransferase"
    /protein_id="AAL16089.1"
    /db_xref="GI:16226163"
    /translation="LEICGSGCYRAVLAAQLVERVPSVRIOLAERAKERLLELRL
    NVFRWGDGEGMPALAPNGIMVTRAAASDVSAALLDOLAPGRLVTPVGSABEQLL
    LIVREPDGGEFRRAHLETVRFPDLQSLF"
    1033..1686
    /gene="mep"
    1033..1686
    /gene="mep"
    /gene="rpos"
    /note="similar to nlpD gene of Pseudomonas aeruginosa"
    /note="transcriptional regulator"
    /codon_start=1
    /transl_table=11
    /product="murein endopeptidase"
    /protein_id="AAL16090.1"
    /db_xref="GI:16226164"
    /translation="MRVGMERELAHNGLIAPVYTHPGOKIQIGPVROAASTARP
    SVPAATPAVTPATSGAAVAAATPVKFDSSKVRPPPTKAPSAQPAVSASPSGMAPA
    EGPLVGRFSSNGSLNKGIDIGQLGPVLAASDGSVYVAGGLRGIGELITIKHSQV
    VSAIGHNRRLLVREGGQVRAQGVIAEMSGTGDVRLKPIREKRGKVPVDIQLBSR"
    1786..2790
    /gene="rpos"
    1786..2790
    /gene="rpos"
    /note="rpos"
    /note="transcriptional regulator"
    /codon_start=1
    /transl_table=11
    /product="stationary phase sigma factor"
    /protein_id="AAL16091.1"
    /db_xref="GI:16226165"
    /translation="MALNITAKAPPEFDVDDVILMDADIVLEDVIDDEPDLPVAKSR
    QANLKQHKYIDYTRALDAQQLYLINEIGFSPLLTPAEVDFPARAOGDPAQRKRMIES
    NLRLLVYKARIARYNRGLSLDLIEBENGLIRAVEKFEDEPRFSTVATWMTROTIE
    RIMNOTRTIRLPIHYVKELNLYLRAARELTOKLDHEPSEAEIATNLEKRVAEVKRL
    GINERTSVDSVLSGPDSDKLTLDVLTFRDPRSDCDLDDDLSDSIDQWLFDLTEKOR
    EYVYRFRGLRGHESCTLEEVGEIGLTRERVRIOYEALKRLREILEKNGLSDALFO"

BASE COUNT      509 a      846 c      959 g      570 t
ORIGIN

alignment_scores:
    Quality:      74.00      Length:      82
    Ratio:        1.510      Gaps:        3
    Percent Similarity: 59.756      Percent Identity: 26.829

alignment_block:
US-10-048-197-2 x AF421351 ..
Align seg 1/1 to: AF421351 from: 1 to: 2884

6 IleArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCys 22
  :::|||||.....:|||||:::|||||.....:
874 GTTCGAGCGTCTGGGGGAGTCTCTGCGCTGTCGGCGGCGTg 923
  22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleG 39
  ||:::|||||:::|||||
924 CGCAGATTCC.....TCCGGAAACAAGCTGCAAGTGGCGG 958
  39 IuGlnThrIleIleLysGlyLysThrAsnLysGlnGlnIleSerSerArg 55
```



```

11:::      :::      111:::      111111
959 AGGCGCAGTCGGTTCCTCCGTCACCGGCGGACATGTCACCCG 1008
56 PhcGlySerAlaSpSerIleSerPheMetIleValIleIleSpheG1 72
1009 GCGGAGTCATCTCGTCATTCGCC.....ATGCGCATATG 1043
72 Y.....HisThrAlaIleLeuAlaPro 79
1044 CTGGGAGTGGCCGACACTGGCCGCCCAATGATCGTCGCC 1089
seq_name: gb_htg:AC067965
seq_documentation_block:
LOCUS AC067965 64165 bp DNA linear HTG 15-JUN-2001
DEFINITION Arabidopsis thaliana chromosome 1 clone T28N5, *** SEQUENCING IN
PROGRESS ***, 11 unordered pieces.
ACCESSION AC067965
VERSION AC067965.2 GI:14456635
KEYWORDS HTG: HTGS_PHASE1.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 64165)
Shim,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C.,
Khan,S., Kim,C., Altafi,H., Bei,O., Chin,C., Chlou,J., Choi,E.,
Conn,L., Conway,A., Gonzales,A., Hansen,N., Howling,B., Koo,T.,
Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharshy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J.,
Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G.,
Federpiel,N.A., Theologis,A. and Ecker,J.R.
Genomic sequence for Arabidopsis thaliana BAC T28N5 from chromosome
1
TITLE
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 64165)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
On Jun 15, 2001 this sequence version replaced gi:7656669.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
4995: contig of 4995 bp in length
4996 19478: contig of 14483 bp in length
19479 21387: contig of 1909 bp in length
21388 23278: contig of 1891 bp in length
23279 24511: contig of 1233 bp in length
24512 26159: contig of 1648 bp in length
26160 29239: contig of 3080 bp in length
29240 32881: contig of 3642 bp in length
32882 37985: contig of 5104 bp in length
37986 60939: contig of 22954 bp in length
60940 64165: contig of 3226 bp in length.
Location/Qualifiers
FEATURES

```

```

source 1. .64165
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="T28N5"
BASE COUNT 20083 a 12418 c 12567 g 19097 t
ORIGIN
alignment_scores:
Quality: 74.00 Length: 67
Ratio: 1.947 Gaps: 5
Percent similarity: 56.716 Percent identity: 34.328
alignment_block:
US-10-048-197-2 x AC067965 ..
Align seg 1/1 to: AC067965 from: 1 to: 64165
41 ThrIleIleLysGlySerThrAsnLysGlnGluIleSerSerArgPheG1 57
:::111:::
56574 TCATCAACGCGACCAACGATCCGACAAA..... 58606
58607 TCACTCGATCAACG.....ATGCGATGCTCCTCTCATTTGGTCAT. 58648
74 hrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIle 90
58649 .....CCAAACAAATGGCTCAAAATACTCCCTCCATTTTA 58684
91 .....SerPheLeu...TrpValLysProTrpArgPr 100
58685 CTGCAATCACACAGTTGAGATCTTTCTTGACTTGTTGAAATTTACTGAC 58734
100 o 100
58735 G 58735
seq_name: gb_htg:SPNED1902
seq_documentation_block:
LOCUS SPNED1902 151947 bp DNA linear HTG 11-JUL-2001
DEFINITION Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
in ordered pieces.
ACCESSION AL449924
VERSION AL449924.1 GI:11545149
KEYWORDS HTG: HTGS_PHASE2.
SOURCE Streptococcus pneumoniae.
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 151947)
AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldera,F., Humbert,Y.,
Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
and Garcia-Bustos,J.F.
Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19F clinical isolate
Microb. Drug Resist. 7 (2), 99-125 (2001)
21353329
2 (bases 1 to 151947)
Dopazo,J., Mendoza,A., Herrero,J., Caldera,F., Polissi,A.,
Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
Garcia-Bustos,J.F.
Direct Submission
JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
TITLE
JOURNAL
COMMENT
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

```

FEATURES
source                                Location/Qualifiers
                                     1. 151947
                                     /organism="Streptococcus pneumoniae"
                                     /serotype="19F"
                                     /db_xref="taxon:1313"
                                     /clone="G54"
BASE COUNT      45278 a 33410 c 27182 g 46003 t      74 others
ORIGIN

alignment_scores:
    quality:      74.00      length:      128
    ratio:        1.121      gaps:         7
    percent similarity: 51.562      percent identity: 29.688

alignment block:
US-10-048-197-2 x SPNEUJ902/rev ..

Align seg 1/1 to reverse of: SPNEUJ902 from: 1 to: 151947

10 1leSerAlaLeuDeuThraL..LeuLeuValThrGlyCysValSerTh 25
      :::::::::::::::::::::
119646 CTAGAGACCTTGAGTACAGCTGATTCGTACAGAGCTCTGTCTAG 119597
      :::::::::::::::::::::
25 r.....GlyAsnValAlaMetGlySerGluGlnAsnGlnGlnThrIleGluG 40
      :::::::::::::::::::::
119596 CATTTTGAAAGAAATGTCAGAAAGATTTTGCATCCGACACAGATCAATGATG 119547
      :::::::::::::::::::::
40 1nhrIleIle.....LysGlyLysThrAsnLysGlnGluIleSerSer 43
      :::::::::::::::::::::
119546 AGGGGATTCCTCGTGTAGCAATATTGGCATTCATTCATCTGTTAGCG 119497
      :::::::::::::::::::::
44 .....LysGlyLysThrAsnLysGlnGluIleSerSer 55
      :::::::::::::::::::::
119496 AGTCGTGTGTGTGTGTAGGAAAGACAAAGAAATGAGTCTATTCTGTGCT 119447
      :::::::::::::::::::::
55 gpheGlySerAlaAspSerIleSerPhe.....MetIleVal 67
      :::::::::::::::::::::
119446 GCATTTTCTCGAAGATACGCTAGGCTGGAGCTGTTATTCCTATGCGCA 119397
      :::::::::::::::::::::
67 aValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTyrGln 83
      :::::::::::::::::::::
119396 TTGTCTTCATTTACGACGACTGGTATATCCTAATCCT..... 119359
      :::::::::::::::::::::
84 GluIleLeuSerLeuIleIleSerPheLeuTyrValLysProTyrArgP 100
      :::::::::::::::::::::
119358 ...CTTTGTCCCTTGTCATTTCTTCTTTATTTCTTCA..... 119323
      :::::::::::::::::::::
119322 .AAGGCCCTTTCACGTTTGTGTACACATCAAG 119290
      :::::::::::::::::::::
seq_name: gb_htg.AC009452

seq_documentation_block:
LOCUS      AC009452      174431 bp      DNA      linear      HTG 30-DEC-2001
DEFINITION Homo sapiens chromosome 17 clone RP11-4F22, map 17, *** SPDUNCING
IN PROGRESS ***, 3 unordered pieces.
ACCESSION  AC009452.17 GI:11560272
VERSION    AC009452.17
KEYWORDS   HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 174431)
            Birren,B., Linton,L., Nusbaum,C. and Lander,E.
            Homo sapiens chromosome 17, clone RP11-4F22
            Unpublished
            2 (bases 1 to 174431)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barnes,N., Beckert,J.R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,

```

```

TITLE
JOURNAL
COMMENT
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 1921
Center clone name: 4_F_22
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 91566: contig of 91566 bp in length
*
* 91567 91666: gap of 100 bp
*
* 91667 121592: contig of 29926 bp in length
*
* 121593 121692: gap of 100 bp
*
* 121693 174431: contig of 52739 bp in length.
*
Location/Qualifiers
1. 174431
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-4F22"
/clone_lib="RCR1-11 Human Male BAC"
BASE COUNT 52366 a 39072 c 35991 g 46800 t 202 others
ORIGIN
alignment_scores:
Quality: 74.00 Length: 90
Ratio: 1.423 Gaps: 2
Percent Similarity: 57.778 Percent Identity: 31.111
alignment_block:
US-10-048-197-2 x AC009452 ..
Align seg 1/1 to: AC009452 from: 1 to: 174431
11 SerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGlyAs 27
||||| ||||| ||| ::|||::||| ::|||:: |||::
63447 TCCCACTACTCTCGGAGACTGACACTGAGAACTGTTAACCACAGAG 63496
27 nValAlaMetLysGlu.GlnAsnGlnGlnThrIleGluInThrIleIle 43
::: ::||| |||::||| |||:: |||
63497 CAGAGGCTTCACGTGACCAAGATCAAGCACTGCTCCAGCCTTGACGA 63546
44 LysGlyLysThr.....AsnLysGlnGlnLui 52
:::|||||

```

```

63547 CAGAGCAAGACCGTCTCAGAGAAAGAAAAAAGAAATG 63596
52 eserSerarphneglySerAlaaspSerIleSerPheMetIleValI 69
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
63597 CACTCACACCTGCACCTTACGACACACTGACGACTGACGATTGTA 63646
63647 TTAAGTATCATCATCCCTCCCTGACGCTCCACAGACAGAAAGCTG 63696
86 leuserleuIleIleSer 91
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
63697 CTCCCTTAACAGCTGCTCG 63714

seq_name: gb_htg:AC099358
seq_documentation_block:
LOCUS AC099358 177495 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus chromosome Rfa clone CH230-34K4, *** SEQUENCING
IN PROGRESS ***, 66 unordered pieces.
ACCESSION AC099358 GI:117943937
VERSION AC099358.2 GI:117943937
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
1 (bases 1 to 177495)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshrocks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burdett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flaggs,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Hulik,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichteberg,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenko,S.,
Oguy,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., RojudoKan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
Slisdon,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,
Stone,H., Sulton,A., Svatek,A., Tabori,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
TITLE JOURNAL
REFERENCE unpublished
AUTHORS 2 (bases 1 to 177495)
TITLE Direct Submission
Direct Submission

```

JOURNAL COMMENT

```

Submitted (10-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:16901764.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GBSX
Center clone name: CH230-34K4
----- Summary Statistics
Assembly program: Phrap; version 0.990329first call to
findhaplolist
Consensus quality: 149775 bases at least Q40
Consensus quality: 157121 bases at least Q30
Consensus quality: 162139 bases at least Q20
Estimated insert size: 143076; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; average-tp estimation
Quality coverage: 2.7x in Q20 bases; sum-of-coverage estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 66 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
5095: contig of 5095 bp in length
5096: gap of unknown length
5195: gap of unknown length
5196: gap of 6188 bp in length
11383: gap of unknown length
11384: gap of 5200 bp in length
16683: gap of unknown length
16783: gap of 5825 bp in length
16784: gap of unknown length
22608: gap of unknown length
22709: gap of 5230 bp in length
27938: gap of unknown length
28039: gap of 4292 bp in length
32330: gap of unknown length
32331: gap of unknown length
32430: gap of 5283 bp in length
32431: gap of unknown length
37813: gap of unknown length
37814: gap of 4379 bp in length
42192: gap of unknown length
42293: gap of unknown length
45821: contig of 3529 bp in length
45822: gap of unknown length
49329: contig of 3408 bp in length
49330: gap of unknown length
49429: gap of unknown length
49430: gap of 3954 bp in length
53383: contig of 3577 bp in length
53484: gap of unknown length
57060: contig of 3577 bp in length
57061: gap of unknown length
57160: gap of 3335 bp in length
60495: contig of 3335 bp in length
57161: gap of unknown length
60595: gap of unknown length
60596: gap of 3437 bp in length
64032: contig of 3437 bp in length
64033: gap of unknown length
64132: gap of unknown length
68040: contig of 3908 bp in length
68041: gap of unknown length
68042: gap of unknown length
71756: contig of 3616 bp in length
71757: gap of unknown length
71856: gap of unknown length
77090: contig of 5234 bp in length
77091: gap of unknown length
77190: gap of unknown length
80409: contig of 3219 bp in length
80410: gap of unknown length
84845: contig of 4336 bp in length
84846: gap of unknown length
84946: gap of unknown length
88506: contig of 3561 bp in length
88507: gap of unknown length
91717: contig of 3111 bp in length

```


of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6.

FEATURES

source

Location/Qualifiers

1..182231

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/clone="RP11-146B14"

/clone_1bp="RPCR-11.1"

7..689

misc_feature

/note="match: GSS: Em:AQ375850"

365..649

repeat_region

/note="AluX repeat: matches 1..286 of consensus"

2945..3711

misc_feature

/note="match: GSS: Em:AQ745052"

3339..3449

repeat_region

/note="L2 repeat: matches 2577..2704 of consensus"

3655..3710

repeat_region

/note="28 copies 2 mer at 71% conserved"

4627..4758

repeat_region

/note="FLAM_C repeat: matches 1..130 of consensus"

5487..5608

repeat_region

/note="MIR repeat: matches 60..173 of consensus"

5775..6071

repeat_region

/note="AluX repeat: matches 1..299 of consensus"

6886..6984

repeat_region

/note="MIR repeat: matches 48..136 of consensus"

7180..7468

repeat_region

/note="AluX repeat: matches 7..299 of consensus"

8427..8541

repeat_region

/note="Alu repeat: matches 22..136 of consensus"

9507..9649

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

10006..10235

repeat_region

/note="MIR repeat: matches 48..188 of consensus"

repeat_region

17112..17420

/note="AluX repeat: matches 1..302 of consensus"

repeat_region

17636..17681

/note="23 copies 2 mer tg 78% conserved"

repeat_region

18345..18526

/note="L1MD1 repeat: matches 6036..6219 of consensus"

misc_feature

18595..19112

/gene="BA146B14.1"

repeat_region

/note="match: GSS: Em:AQ548118"

19305..19618

/note="Alu repeat: matches 1..312 of consensus"

misc_feature

complement(19955..20234)

misc_feature

19974..20146

/gene="BA146B14.1"

repeat_region

/note="match: STS: Em:G15769"

20136..20513

/note="L2 repeat: matches 2310..2718 of consensus"

repeat_region

21215..21508

/note="Alu repeat: matches 1..292 of consensus"

repeat_region

22510..22813

/note="AluB repeat: matches 3..297 of consensus"

repeat_region

23184..23313

/note="FLAM_A repeat: matches 1..129 of consensus"

repeat_region

25395..25719

/note="MERIB repeat: matches 1..331 of consensus"

repeat_region

26758..26797

/note="20 copies 2 mer ac 90% conserved"

repeat_region

27107..27399

/note="Alu repeat: matches 1..293 of consensus"

misc_feature

complement(27309..27637)

repeat_region

27422..27556

/note="L1PB3 repeat: matches 6010..6141 of consensus"

misc_feature

complement(28428..28893)

repeat_region

29079..29160

/note="match: GSS: Em:AQ691327"

repeat_region

29947..30240

/note="AluX repeat: matches 1..294 of consensus"

repeat_region

32144..32184

/note="MIR repeat: matches 70..111 of consensus"

repeat_region

33021..33313

/note="AluX repeat: matches 7..299 of consensus"

repeat_region

33452..33627

/note="L1M47 repeat: matches 6093..6279 of consensus"

repeat_region

33871..34002

/note="MIR repeat: matches 8..140 of consensus"

repeat_region

34003..34292

/note="AluX repeat: matches 1..285 of consensus"

repeat_region

36003..36280

/note="L1R22 repeat: matches 188..509 of consensus"

repeat_region

36281..36579

/note="AluX repeat: matches 1..300 of consensus"

repeat_region

36580..36780

/note="L1R22 repeat: matches 1..188 of consensus"

repeat_region

38781..39023

/note="Charliel repeat: matches 2522..2761 of consensus"

repeat_region

39032..40398

/note="L1M3e repeat: matches -878..1358 of consensus"

repeat_region

40400..40895

/note="Charliel repeat: matches 4..548 of consensus"

repeat_region

40913..41316

/note="TIGER2 repeat: matches 1..450 of consensus"

repeat_region

41889..42199

/note="AluX repeat: matches 1..311 of consensus"

repeat_region

43198..43468

/note="Alu repeat: matches 38..298 of consensus"

misc_feature

complement(45910..46385)

repeat_region

46043..46076

/note="match: GSS: Em:AQ608877"

misc_feature

46392..46807

/note="17 copies 2 mer ac 100% conserved"

```

/gene="bA146B14.1"
/!note="match: GSS: Em:AQ175358"
repeat_region 47247..47549
/!note="AluY repeat: matches 1..303 of consensus"
repeat_region 47550..47848
/!note="AluX repeat: matches 1..300 of consensus"
repeat_region 48060..48358
/!note="AluJ repeat: matches 2..299 of consensus"
repeat_region 49173..49484
/!note="MER2 repeat: matches 3..346 of consensus"
repeat_region 51706..52084
/!note="L1MA2 repeat: matches 5937..6305 of consensus"
misc_feature 52075..52416

alignment_scores:
  Quality: 74.00      Length: 99
  Ratio: 1.396        Gaps: 5
  Percent Similarity: 53.535      Percent Identity: 29.293

alignment_block:
US-10-048-197-2 x AL137020 ..

Align seg 1/1 to: AL137020 from: 1 to: 182231

22 CysValSerThr.....GlyAsnValAlaMetLysGluGlnAsnGlnG1 36
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166688 TGCTAAACACAGATGTAGACGAGTGTGACATGCTCTGAGCAGCAA 166737
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
36 nThrlleGluGlnThrllelleGlyGlyThrAsnLysGlnGluLies 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166738 TACAAATACCCAGGAAGATT.....GGAAGAGCTAAGAAAAAGCCATTC 166781
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53 erSerArpHeglySerAlaSerIleSerPheMetIleValAlaIle 69
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166782 CCAGAAACTT.....TTCAGCTATGGCATCTTGGGAAGC 166816
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
70 LysPheGlyHisThrAlaIleLeuAla..... 78
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166817 AATTTTGATCTTCGTCGTCATCTTGACCTTGAGTGTCTGAGATTTC 166866
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 .....ProAsnArgTTPGlnGlnIleLeu.SerLeuIleIleSer 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166867 GCCCAAGTGGCCAGACAGACAGATCTCTCATGTCAGTGGCCCACT 166916
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92 PheLeuTTPValLysProTyrArpProLysAsnLeuSerPheTyr 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166917 CAGCTTTGGGAAGTCAATCCATCCAGCCCATGATGATTAT 166961
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

seq_name: gb_htg:AC099386

seq_documentation_block:
LOCUS AC099386 217121 bp DNA linear HTG 21-DEC-2001
DEFINITION Rattus norvegicus chromosome Rf4 clone CH230-163F6, WORKING DRAFT
SEQUENCE 35 unordered pieces.
AC099386.4 GI:17974833
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 217121)
1 (bases 1 to 217121)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barberia,J.,
Benton,T., Bimagne,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhey,C.,
Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

```

TITLE JOURNAL REFERENCE AUTHORS JOURNAL COMMENT

```

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsí,F., Howard,S., Huber,J., Huyls,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W.,
Loudseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,G., Nguyen,N., Nickerson,E., Nwokemwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshari,N.,
Slisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,A., Weczyk,R., Wooden,S.,
Watlington,S., Williams,G., Williamson,R., Wleczky,R., Wooden,S.,
Morley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 217121)
Morley,K.C.
Submitted (10-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 21, 2001 this sequence version replaced gi:17062973.

--- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
--- Project Information
Center project name: GKPA
Center clone name: CH230-163F6
--- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 199000 bases at least Q40
Consensus quality: 202571 bases at least Q30
Consensus quality: 205231 bases at least Q20
Estimated insert size: 199252; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; gapwise-tp estimation
Quality coverage: 4.5x in Q20 bases; sum-of-contigs estimation
---
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
17031 17030: contig of 17030 bp in length
17031 17130: gap of unknown length
17131 34262: contig of 17132 bp in length
34262 34362: gap of unknown length
34363 49178: contig of 14816 bp in length
49179 49278: gap of unknown length
49279 64140: contig of 14862 bp in length

```


THIS PAGE BLANK (USPTO)

```

27  nvalalameltyglnghlnbsnclnfnhrilleglnghlnthrllelel 44
    :::: ||:::||||: ||:::||||:
585  ....TtATTGACACGAATGCAATTAAGAAATACAGACATTATATAC 543
    :::: ||:::||||: ||:::||||:
44  yscglysthraslnlysglnghlnlserSerargPheglySerAlaasp 60
    :::: |||: ||:::||||: ||:::||||:
542  AA...AAATCTCATAATTCAGACTTCTCTTCAAAATTCACAGGACGAAT 496
    :::: |||: ||:::||||: ||:::||||:
61  SerIleSerPheMetIleValValIleIeIysPheGlyHisThrAlaIle 77
    |||: :::: ||:::||||: ||:::||||: ||
495  GTATTATTGATTGGACAGAGTTATTATTAGACAGTCTCGAATGTATACATT 446
    |||: :::: ||:::||||: ||:::||||: ||
77  uAlaPcOAsnAArgTIpGlnGlnIleIeuSerIleuIleIleSerPheLeuT 94
    |||: :::: |||: ||:::||||: ||:::||||: ||:::||||:
445  ATTAAACACAGCTGGGAACCTGTATTAGATGTAAATAAAATTTAAAAA 396
    |||: :::: |||: ||:::||||: ||:::||||: ||:::||||:
94  rPValIysProTyrArgPProIysAsn 102
    |||: :::: |||: ||:::||||: ||:::||||: ||:::||||:
395  TTATAAATCAGCAGCAAGGATATTTAAT 370
    |||: :::: |||: ||:::||||: ||:::||||: ||:::||||:
seq_name: gb_gss:BH150303

```

seq_documentation_block: 923 bp DNA linear GSS 27-AUG-2001

LOCUS BH150303

DEFINITION ENRP129TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.

ACCESSION BH150303

VERSION BH150303.1 GI:15311742

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica

REFERENCE 1 (bases 1 to 923)

AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.

TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HM1:IMSS sheared DNA library (2001)

COMMENT Unpublished (2001)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3563
Email: bj1ofutsel@igf.org
Clones are derived from the *Entamoeba histolytica* HM1:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 19
High quality sequence stop: 159.

FEATURES
source

```

/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="vector: pROSI; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
Whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaundin and B.
Barell, Oxford University Press, 1993)."

```

```

alignment_scores:      84.00      Length:      92
                       Quality:    Gaps:      2
                       Ratio:      1.556
Percent Similarity:    58.696      Percent Identity: 28.261

alignment_block:
US-10-048-197-2 x BH150303/rev  ..

```

Align seg 1/1 to reverse of: BH150303 from: 1 to: 923

[illegible]

seq_name: gb_est1:AW352637

seq. documentation block:	617 bp	mRNA	linear	EST 02-FEB-2000
LOCUS	AM352637			
DEFINITION	660033id06.x1 660 - Mixed stages of anther and pollen Zea mays CDNA,			
ACCESSION	AM352637 mRNA sequence.			
VERSION	AM352637.1	GI:6851627		
KEYWORDS	EST.			
SOURCE	Zea mays.			
ORGANISM	Zea mays			

REFERENCE
AUTHORS
TITLE

Maize ESTs from various cDNA libraries sequenced at Stanford
Walbot, V.
1 (bases 1 to 617)
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Kerkyota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V

Department of Biological Sciences
Stanford University
855 California Ave., Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660003 row: D column: 06.
Location/Qualifiers

Source

```

/organism="Zea mays"
/cultivar="Oh1043"
/db xref="taxon:4577"
/clone_lib="660 - Mixed stages of anther and pollen"
/rissue_type="Whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="XLOLR"
/note="Organ: anthers; Vector: lambda Zap; Site_1: EcoRI
site_2: XhoI; anther and pollen cDNA library.
Directionally sequenced with 5' end at the EcoRI site.

```

Created by Amie Franklin.
 BASE COUNT 173 a 148 c 123 g 173 t
 ORIGIN

alignment_scores:
 Quality: 82.50 Length: 101
 Ratio: 1.587 Gaps: 4
 Percent Similarity: 51.485 Percent Identity: 25.743

alignment_block:
 US-10-048-197-2 x AM352637/rev ..

Align seg 1/1 to reverse of: AM352637 from: 1 to: 617

```

7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCysVala 23
   ::::::::::::::::::::|||:::|||||
379 AAGGCTACAGTCAAGTTCACAGAAAGATCCTTCATGGGT...GT 333
23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
   ::::::::::::::::::::|||:::|||||
332 CGCGGTGTGTACCTGTCTATGGAGGAGAACAGATCCACAGAACATCC 283
39 lGlnThrIleIleLysGlyLysThrAsnLysGlnGlnIleSerSerArg 55
   ::::::::::::::::::::|||:::|||||
282 AGATG..... 278
56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG1 72
   ::::::::::::::::::::|||:::|||||
277 .....ACGATCAACTTCCTCGTCTCTCTCTGGAAG..... 248
72 yHstHrAlaIleLeuAlaProAsnArgTPrGlnGlnIleLeuSerLeu 89
   ::::::::::::::::::::|||:::|||||
247 .....AAGAACTGGCAGAAATGTAAGATGTCCT 220
89 lIleSerPheLeuTrpValIleSerProTyrArgProLysAsnLeuSerPhe 105
   ::::::::::::::::::::|||:::|||||
219 ACATCAAGACACCATCGGGAACCATTCACAGATTAAAAAATAATGTTT 170
106 Tyr 106
   ::
169 TTT 167

```

seq_name: gb_est1:BB840173

seq_documentation_block:

LOCUS BB840173 459 bp mRNA linear EST 21-NOV-2001
 DEFINITION BB840173 RIKEN full-length enriched, 12 days embryo whole body Mus
 musculus cDNA clone E97005N21 5', mRNA sequence.

ACCESSION BB840173
 VERSION BB840173.1 GI:17040904

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus.

REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 459)

Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
 Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,
 Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T.,
 Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
 Saito,K., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
 Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,
 A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toyota,T.,
 Wataniki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 2001)

Unpublished (2001)

CONTACT: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-resesc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,
 M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,
 Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

FEATURES
 source Location/Qualifiers

1..459
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="E97005N21"
 /clone_lib="RIKEN full-length enriched, 12 days embryo
 whole body"
 /tissue_type="whole body"
 /dev_stage="12 days embryo"
 BASE COUNT 126 a 70 c 126 g 137 t
 ORIGIN

alignment_scores:
 Quality: 81.00 Length: 137
 Ratio: 1.191 Gaps: 4
 Percent Similarity: 49.635 Percent Identity: 22.628

alignment_block:

US-10-048-197-2 x BB840173/rev ..

Align seg 1/1 to reverse of: BB840173 from: 1 to: 459

```

4 HstHrAlaIleLeuThrIleSerAlaLeuThrAlaLeu.....Le 18
   ::::::::::::::::::::|||:::|||||
455 CATACTTCTTGTCCTCTATAGGATTTCTGACCTGTCTTGACGCT 406
18 uValThrGlyCysValSerThrGlyAsnValAlaMetLysGlnGlnang 35
   ::::::::::::::::::::|||:::|||||
405 CTACACACAGTTTCAACAAGTACAGAGTCCATATATACAGAACAAACA 356
35 lGlnThrIleGlnGlnThrIleIleLysGlyLysThrAsnLysGlnGlu 51
   ::::::::::::::::::::|||:::|||||
355 AGACCAAACTATCAGACTACCCACCAAAAGGTTACGGGGAAGAAAAAC 306
52 lIleSerArgPheGlySerAlaAspSerIleSerPheMetIleVala 68
   ::::::::::::::::::::|||:::|||||
305 ATACTACACAGCAGGATCTATTGATCAGTACGCTGAGGTATCATAGT 256
68 lIle..... 69
255 TCTTAACCTGTGCCCAAAACAGACACTTAAGTGTCCCAAAAGCCAA 206
70 .....LysPhe 71
205 GGAACACTTAAGTGTCCCAAAACGGGGAACATTAAGTGTCCCAAAAGCC 156
72 GlyHstHrAlaIleLeuAlaProAsnArgTPrGlnGlnIleLeuSerL 88

```

```

155 GGAACCTATCGTTACTTGTGCTCCTTCCTGGCAA.....C 118      ||||| |
105 Phenyleu 107          :::::||||| :||| |
67 TACTTGTC 59          :::::|::| |
                                :::::|::| |
seq_name: gb_est2:BE575116
seq_documentation_block:
LOCUS       BE575116                      457 bp     mRNA     linear   EST_15-AUG-2000
DEFINITION  946087A11.x2 946 - tassal primordialium prepared by Schmidt lab Zea
ACCESSION   BE575116
VERSION     BE575116.1  GI:9824809
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays.
REFERENCE   Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS     Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
TITLE        clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 457)
AUTHORS     Walbot,V.
TITLE        Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL     Unpublished (1999)
COMMENT     Contact: Walbot, V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
FEATURES             Location/Qualifiers
     source           1..457
                     /organism="Zea mays"
                     /cultivar="OH43"
                     /db_xref="taxon:4577"
                     /clone_lib="946 - tassal primordialium prepared by Schmidt
                     lab"
                     /tissue-type="tassels"
                     /dev_stage="just after the transition from vegetative to
                     inflorescence development"
                     /lab_host="XLOLR"
                     /note="Organ: tassels; Vector: HybridAP; Site_1: ECOR1;
                     Site_2: XhoI; George Chuck dissected Immature tassels
                     between lmm and 3mm. Sharon Stanfield prepared the cDNA
                     library in HybridAP. Sample insert size range was 350 bp
                     to 3 Kb with a 1 Kb average."
BASE COUNT      114 a                118 c                107 g                118 t
ORIGIN
alignment_scores:
Quality:         76.50              Length:          94
Ratio:           1.628              Gaps:            4
Percent Similarity: 50.000          Percent identity: 25.532
alignment_block:
US-10-048-197-2 x BE575116/rev ..
Align seg 1/1 to reverse of: BE575116 from: 1 to: 457
7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGLyCyVa 23
::: ||::::: ||| ::||| |||
288 AAGCAACCGTCACACTCCAGCTCAAGAAGCGCTTTGCATGGCC..GT 242
23 IserThrGlyAsnValAlaMetLysGlu...GlnAsnGInGInThrIleG 39
```

```

|||||.....|||
241 CGCGCTGGCACTTGCCATGAGGAGAAGCAGATCCACGACAATCC      192
39 InGlAnThrIleLeuYsgLYstHrAsnLysInGluLeuSerArg 55
   ::::
191 AGATG.....                                     187
56 PheGLySeRAlaSpErTLeSeRPHeMeTiLeValIleLysPheGI 72
   |||||.....|
186 .....AGCGTAACCTTCGTGTTCCATTATGAAG..... 157
72 yHISThrAlAileuAlAProASnaRgrTPngInGuileuSerLeuI 89
156 .....AAGAAClGGCACGATGAGATGCCTCT 129
89 lEIlSeRPHeLeUTrpYAlLySProTYrARg 99
   ::::|
128 ACGTCAGAGCACCATGCGGAAGCCATACAG 97

seq_name: gb_estl:A1833467

seq_documentation_block:
LOCUS       A1833467              547 bp    mRNA          linear     EST-02-FEB-2000
DEFINITION  605089G04.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
ACCESSION   A1833467
VERSION     A1833467.1 GI:5455777
KEYWORDS    EST.
SOURCE      Zea mays.
            Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 547)
AUTHORS     Walbot,V
TITLE        Maize ESTs from various cDNA libraries sequenced at Stanford
             University
JOURNAL     Unpublished (1999)
COMMENT     Contact: Walbot V
             Department of Biological Sciences
             Stanford University
             855 California Ave, Palo Alto, CA 94304, USA
             Tel: 650 723 2227
             Fax: 650 725 8221
             Email: walbot@stanford.edu
             Plate: 605089 row: G column: 04.
FEATURES             location/Qualifiers
             source
               1..547
               /organism="Zea mays"
               /cultivar="Ohio43"
               /db_xref="taxon:4577"
               /clone_lib="605 - Endosperm cDNA library from schmidt lab"
               /tissue_type="nucellar, embryo, and endosperm"
               /dev_stage="10-14 days post-pollination"
               /lab_host="DH5(alpha)"
               /note="Organ: kernel; Vector: pAD-GAL4-2'; Site_1: ECORT;
               Site_2: Xhol; Kernel endosperm cDNA library from Schmidt
               lab"
BASE COUNT      152 a      132 c      123 g      140 t
ORIGIN
alignment_scores:
  Quality:      76.50      Length:      94
  Ratio:        1.628      Gaps:      4
  Percent Similarity: 50.000      Percent Identity: 25.532
alignment_block:
US-10-048-197-2 x A1833467/rev ..
Align seg 1/1 to reverse of: A1833467 from: 1 to: 547
7 ArgLeuTrIleSeRAlaleuLeuThraLaleuLeuValIrHgLYcysVa 23

```

```

386  AGGCAACCGTCAGTTCAGCTCAAGACGTCCTTGCATGGCC...GT 340
      |||
      |||
      |||
23  lserThrGlySerValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
      |||
      |||
      |||
339  CCCCTCCGCACTTCGGCATGAGAGACAGATCCAGACATCC 290
      |||
      |||
      |||
39  lncGlnThrIleLeuLysGlyThrAsnLysGlnGlnLysSerArg 55
      |||
      |||
      |||
288  AGATG..... 285
56  PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG 72
      |||
      |||
      |||
284  .....ACCGTGAACCTTCCTGTTCCCTTTGACG..... 255
72  yHisThrAlaIleLeuAlaProAsnArgTrpGlnGlnIleLeuSerLeu 89
      |||
      |||
      |||
254  .....AGCAACTGCGCAGATGTGAGATGCTTT 227
89  lclleSerPheLeuTrpValLysProTyrArg 99
      |||
      |||
      |||
226  ACGTCAAGACGACCATGGGAAACCATACAG 195
seq_name: gb_est2:T41696
seq_documentation_block:
LOCUS T41696 607 bp mRNA linear EST 07-AUG-1995
DEFINITION 102277 Lambda-PRL2 Arabidopsis thaliana cDNA clone 64B87, mRNA
sequence.
ACCESSION T41696
VERSION T41696.1 GI:931332
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Rosidae; eucotsids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 607)
Newman,T., deBruin,J.F.J., Green,P., Keegstra,K., Kende,H., McIntosh
,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,N., Retzel
,E. and Somerville,C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)
95148729
On Jan 24, 1995 this sequence version replaced gi:634284.
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313c@clm.c1.msu.edu
Seq primer: T7 dye primer.
FEATURES
SOURCE
1. 607
/organism="Arabidopsis thaliana"
/strain="var Columbia"
/db_xref="taxon:3702"
/clone="64B87"
/note="Vector: lambda zip-lox; site_1: Sal; site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRL's lambda zip-lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dT primed cDNA."
BASE COUNT 139 a 128 c 146 g 164 t 30 others

```

```

ORIGIN
alignment_scores:
  Quality: 76.50 Length: 95
  Ratio: 1.319 Gaps: 5
Percent Similarity: 61.053 Percent Identity: 25.263
alignment_block:
US-10-048-197-2 x T41696 ..
Align seg 1/1 to: T41696 from: 1 to: 607
3 AsnHisHisIleArgLeuThrIleSerAlaLeuThr..... 15
|||||
|||||
176 AAcAGAAAGATGATCGGTATCCGTCGCCGCCGCAACAGCAACGACNNNT 225
|||||
16 .AlaLeuLeuValThrGlyCysValSer...ThrGlySerValAla.... 29
|||||
226 CACCGTCTCGTTACCGGTGCGGNTGAAAGAACAGACAAATGTGTATA 275
|||||
30 ....MetLysGlnGlnAsnGlnGlnThrIleGlnGlnThrIleLys 44
|||||
276 AGAAATTGAAGAGAGAGCTGCGACAGTTCGTGCAAGGCGTTAGTACAG 325
|||||
45 GlyLysThrAsnLysGlnGlnLysSerArgPheGlySer..... 58
|||||
326 ACAAAGAGAGATGAGAGAAATCAATGAGAGATGAAATGATTCATT 375
|||||
59 .....AlaAspSerIleSerPheMetIleValIleLys.... 70
|||||
376 GCGNGATATTAGNGATACATGCATCAATNTGCTCCGCTGTTGAAGGG 425
|||||
71 .....PheGlyHisThrAlaIleLeuAla 78
|||||
426 NTYTNAATGCTTTTGGTCAATCTTACCTTAGCG 460
seq_name: gb_est2:BM412909
seq_documentation_block:
LOCUS BM412909 745 bp mRNA linear EST 22-JAN-2002
DEFINITION EST587236 tomato breaker fruit Lycopersicon esculentum cDNA clone
CLEG61H17 5' end, mRNA sequence.
ACCESSION BM412909
VERSION BM412909.1 GI:18264539
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 745)
Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai
,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Roming
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished (2002)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.
FEATURES
SOURCE
1. 745
/organism="Lycopersicon esculentum"
/cultivar="T496"
/db_xref="taxon:4081"
/clone="CLEG61H17"
/clone_lib="tomato breaker fruit"

```

```

/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/notes="Vector: pBluescriptSKmCUDapt; Site.1: EcoRI;
Site.2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."

BASE COUNT      211 a      123 c      178 g      233 t
ORIGIN

alignment_scores:
  Quality:      76.50      Length:      89
  Ratio:        1.443      Gaps:      4
  Percent Similarity: 59.551      Percent Identity: 29.213

alignment_block:
US-10-048-197-2 x BM412909 ..

Align seg 1/1 to: BM412909 from: 1 to: 745

10  TleseralaleuleuthralaleuleuValThrglyCysValSerThrG1 26
    ::::::::::::::::::::
36  GTTCTGCGCCCAAGTAATCTGTGAAGTTCTTTGTGAAGAGTGCA 85
    ::::::::::::::::::::
26  yasnValAlaMetLysGluGlnAsnGlnGlnThrIleGlnThrIleI 43
    ::::::::::::::::::::
86  GAGATTGGTATGTAAAGCAACA.....GCAACTGATG 117
    ::::::::::::::::::::
43  IelysGlyLysThrAsnLysGlnGlnLysSerSerArgPheGlySerAla 59
    ::::::::::::::::::::
118  TTAATCGTAAAGAGACCAAGATGAA.....TCTTTGGTGAGGCT 158
    ::::::::::::::::::::
60  ..AspSerLysSerPheMetLysValAlaLysPheGlyHisThrAl 75
    ::::::::::::::::::::
159  CCTGATGAAATGACA.....TGCTGTATGAAGTTGGGGTTCATC 199
    ::::::::::::::::::::
75  aTleuAlaProAsnArgTyrGlnGlnLysLeuSerLeuLysSerP 92
    ::::::::::::::::::::
200  AGTAGCCTCAGCGCCGAAGATGAGAGATGAGTATCTTATCTTACT 249
    ::::::::::::::::::::
92  heLeuTrpValLysPro 97
    ::::::::::::::::::::
250  TCTTAGAAGAGAGAGACCT 266

seq_name: gb_est1:AI068415
seq_documentation_block:
LOCUS      AI068415      918 bp      mRNA      linear      EST 09-DEC-1999
DEFINITION  mgae0002bA08f Magnaporthe grisea Appressorium Stage cDNA library
ACCESSION  AI068415
VERSION    AI068415.1 GI:3391390
KEYWORDS   EST.
SOURCE     Magnaporthe grisea.
ORGANISM   Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
REFERENCE  1 (bases 1 to 918)
            Choi, W., Fang, F., Sasinowski, M., Wang, R. and Dean, R.A.
            Expressed sequence characterization during appressorium formation
            in rice blast fungus, Magnaporthe grisea
            Unpublished (1998)
            Contact: Dean, R.A.
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
            Tel: 864 656 5737
            Fax: 864 656 4293
            Email: rdean@clemson.edu

```

```

Seq primer: T3 primer (AATTAACCTCCTCAATAAGG)
High quality sequence stop: 331.
Location/Qualifiers
1..918
/organism="Magnaporthe grisea"
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgae0002bA08f"
/clone_1lb="Magnaporthe grisea Appressorium Stage cDNA
library"
/dev_stage="Germinated conidia on appressorium-inductive
surface"
/notes="Vector: pBluescriptII SK(+) Vector; Site.1: EcoRI;
Site.2: XhoI; the appressorium formation-specific cDNA
library was constructed from conidia germinated for 5-8
hr on an inductive surface. The library has an average
insert size of 1.5 kbp."

BASE COUNT      241 a      237 c      227 g      201 t      12 others
ORIGIN

alignment_scores:
  Quality:      76.50      Length:      113
  Ratio:        1.594      Gaps:      3
  Percent Similarity: 42.478      Percent Identity: 22.124

alignment_block:
US-10-048-197-2 x AI068415 ..

Align seg 1/1 to: AI068415 from: 1 to: 918

7  ArgLeuThrIleSerAlaLeuLeuThralaleuleuValThrglyCysVa 23
    ||| ||||| ::::::::::::::::::::
107  AGAACCACTATACCATCTGTCTGTACTCAACAATATACGATGGCTCGT 156
    ::::::::::::::::::::
23  IserThr.....
157  TTCACCCAGACACAGCATTTGCCAGAAACTACTGCGCCAGTTCCA 206
    ::::::::::::::::::::
26  .....GlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleLeu 39
    ||| ::::::::::::::::::::
207  ACAACACAGAGAACACGCTATGAGCAGAGAACGACGACACTATCATC 256
    ::::::::::::::::::::
40  GlnThrIleLysGly.....
257  GAGCTGTTCGCTGCTGTGCTGCTGCTCATCGTCTGCTGTCATG 306
    ::::::::::::::::::::
46  .....LysThrAsnLysGlnLys 53
    ::::::::::::::::::::
307  TCTGTTGCTTGGCGCTGTATGCTGTCCGACAGAACACGAGAGGCTG 356
    ::::::::::::::::::::
53  erSerArgPheGlySerAlaAspSerLysSerPheMetLysValAla 69
    ::::::::::::::::::::
357  ACAACTCTTTGGAACTCTCTCCCTCA.....
70  LysPheGlyHisThrAlaLysAlaProAsnArgTyr 82
    ::::::::::::::::::::
383  .....ACCTCCGTGAGAAAGCCGGAACCGGTGG 409

seq_name: gb_est1:AW933759
seq_documentation_block:
LOCUS      AW933759      537 bp      mRNA      linear      EST 18-MAY-2001
DEFINITION  EST359602 tomato fruit mature green, TAMU Lycopersicon esculentum
ACCESSION  AW933759
VERSION    AW933759.1 GI:8109160
KEYWORDS   EST.
SOURCE     tomato.
ORGANISM   Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

```

REFERENCE 1 (bases 1 to 537)
 AUTHORS Alcalá, J., Vrebalov, J., White, R., Matern, A.L., Holt, I.E., Liang, F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Rønning, C.M., Fraser, C.M., Matlin, G.B., Tanksley, S.D. and Giovannoni, J.
 TITLE Generation of ESTs from tomato fruit tissue
 JOURNAL Unpublished (1999)
 COMMENT Contact: CUGI
 clemson university Genomics Institute
 clemson university
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES
 source
 1..537
 Location/Qualifiers
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="CLEF56018"
 /clone_lib="tomato fruit mature green, TAMU"
 /tissue_type="fruit pericarp"
 /dev_stage="mature green (3-5 days pre-ripening)"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; CLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp"

BASE COUNT 160 a 89 c 114 g 174 t
 ORIGIN

alignment_scores:
 Quality: 76.00 Length: 79
 Ratio: 1.520 Gaps: 3
 Percent Similarity: 63.291 Percent Identity: 26.582

alignment_block:
 US-10-048-197-2 x AW93759/rev ..

Align seg 1/1 to reverse of: AW93759 from: 1 to: 537

```

40 GlnThrIleIleIlys...GlyLysThrAsnLysGlnIleIleSerSerAr 55
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
479 CAACCGCTACTGAAATGTCATCGTGGAAATGTCGACGAACTCCACGATCG 430
55 gPheGlySerAlaAspSerIleSerPhe.....MetIleValy 68
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
429 ATTTCTTCACAGTCGAAATACCTCATTTCCCTTCTTCTGCTCTCTG 380
68 AlIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnIle 84
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
379 TAGTCACACTTTTCACCTTCTAGCTATCCCTCGTCAAGGCGCAAG 330
85 .....IleLeuSerIleIleIleSerPheLeuTrpValy 96
339 TAGCCTCTTTTAACTTTGATCCAACTGTTTCTTTTTCACAACTTCTCA 280
96 sProTyrArgProLysAsnLeuSerPheTyrLeuThr 108
| : : : : | | | : : : : | | | : : : : | | | : : : : |
279 GGAGTTTAGAAACATACCTTTTACTATATATCA 243

```

seq_name: gb_esc12:Bi679657

seq_documentation_block:
 LOCUS Bi679657 370 bp mRNA linear EST 17-SEP-2001
 DEFINITION 949031A01.y3 949 - Juvenile leaf and shoot cDNA from Steve Moose
 ACCESSION Zea mays cDNA, mRNA sequence.
 VERSION Bi679657
 KEYWORDS Bi679657.1 GI:15632564
 EST.
 SOURCE Zea mays.

ORGANISM Zea mays
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae; PACC
 clade: Panicoideae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 370)
 AUTHORS Walbot, V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2327
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 949031 row: A column: 01.

FEATURES
 source
 1..370
 Location/Qualifiers
 /organism="Zea mays"
 /cultivar="M64A"
 /db_xref="taxon:4577"
 /clone_lib="949 - Juvenile leaf and shoot cDNA from Steve Moose"
 /tissue_type="Immature leaf primordium and vegetative meristem"
 /dev_stage="4 stages from 3-13 days after imbibing"
 /lab_host="E. coli XL0LR"
 /note="Organ: Juvenile vegetative shoots; Vector: PAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts of total RNA by weight from 4 tissue sources (see below) were pooled, polyA+ RNA isolated, and cDNA synthesized for EcoRI (5') and XhoI (3') directional cloning into Lambda Hybridap vector from Stratagene. Tissue Sources: 1. Whole shoots 3 days after sowing/imbibing in wet soil. 2. Basal 1.5 cm shoots 6 days after sowing - includes yellow portions of developing leaves 1-5, primordia from 6-8, and the vegetative apex. 3. Non-green portions of developing leaves 4-5 and the vegetative apex, including adult leaf primordia, 9 days after sowing. 4. Partially expanded and greening leaves 4-5 at 13 days after sowing."

BASE COUNT 91 a 97 c 97 g 85 t
 ORIGIN

alignment_scores:
 Quality: 75.50 Length: 94
 Ratio: 1.606 Gaps: 4
 Percent Similarity: 50.000 Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x Bi679657 ..

Align seg 1/1 to: Bi679657 from: 1 to: 370

```

7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCysVa 23
: : : : : | | | : : : : | | | : : : : | | | : : : : |
130 AAGGCAACCGTCAAATTCACGCTCAAGAGGCTCTTGCAAGGCG...GT 176
23 IserThrGlyAsnValAlaMetLysGlu...GlnAsnGlnThrIleG 39
| : : : : | | | : : : : | | | : : : : | | | : : : : |
177 CGCCGTCGCAACTGCGCATGAGAGACAGACAGATCCAGAACATCC 226
39 IuGlnThrIleIleLysGlyLysThrAsnLysGlnIleIleSerSerArg 55
: : : : : | | | : : : : | | | : : : : | | | : : : : |
227 AGATG..... 231
56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG 72
| : : : : | | | : : : : | | | : : : : | | | : : : : |
232 .....AGCGTGAACCTTCTGTTTCCCTTCGAAAG..... 261
72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnIleIleLeuSerLeu 89
: : : : : | | | : : : : | | | : : : : | | | : : : : |

```

```

262 .....AAGACAGCGCAGATGTCGATGCTCT 289
89 le1leSerPheLeuTrpValLysProTyrArg 99
      :::::
      |||||
DEFINITION 61406OH04.x1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
      mRNA sequence.
ACCESSION  AM017581
VERSION    AM017581
KEYWORDS   EST.
SOURCE     Zea mays.
ORGANISM   Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoidae; Andropogoneae; Zea.
            1 (bases 1 to 380)
REFERENCE  Walbot, V.
            Maize ESTs from various cDNA libraries sequenced at Stanford
            University
AUTHORS    Unpublished (1999)
JOURNAL    Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614060 row: H column: 04.
FEATURES
source     Location/Qualifiers
            1..380
            /organism="Zea mays"
            /cultivar="W23"
            /db_xref="taxon:4577"
            /clone_lib="614 - root cDNA library from Walbot Lab"
            /tissue_type="root"
            /dev_stage="3-4 days old"
            /lab_host="X10LR"
            /note="Organ: root; Vector: pBluescriptII SK+; Site:1:
            EcorI; Site:2: XhoI; 3-4 days old root tissue from Walbot
            Lab (Lm)"
BASE COUNT 103 a 101 c 80 g 96 t
ORIGIN
alignment_scores:
      Quality: 75.50      Length: 94
      Ratio: 1.606      Gaps: 4
Percent Similarity: 50.000      Percent Identity: 25.532
alignment_block:
US-10-048-197-2 x AM017581/rev ..
Align seg 1/1 to reverse of: AM017581 from: 1 to: 380
7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
      :::::
      |||||
353 AAGGCAACCGTCMAAGTCCAGCTCAGCAGAGAAGGTCCTTGCATGGGC...GT 307
23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
      l:::::
      |||||
306 CGCCGCGCGCAACTTGCGATGGAGGAGCAAGCATCCACAGCAAAATCC 257
39 lGlnIleThrIleIleLysGlyLysThrAsnLysGlnLulSerSerArg 55
      :::::
256 AAGATG..... 252
56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheGl 72
      |||||

```

```

251 .....AGCGTGAACCTCCTCTGTTTCCCTCTAGAG..... 222
72 yH1sthrAa1leuAlaProAsnAr7TrpGlnGlu1leuSerLeuI 89
221 .....AAGAACGCGCAGATGTGATGCGCTCT 194
89 le1leSerpheLeuTrpValLysProTrpArg 99
193 ACCTCAAGAGACACCATGGGAGACCATACAGC 162
seq_name: gb_est2:B1698181

seq_documentation_block:
LOCUS B1698181 450 bp mRNA linear EST 18-SEP-2001
DEFINITION 949031A01.y4 949 - Juvenile leaf and shoot cDNA from Steve Moose
ACCESSION B1698181
VERSION B1698181.1 GI:15660810
KEYWORDS EST.
SOURCE
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 450)
Walbot.V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949031 row: A column: 01.
FEATURES
Source
1..450
Location/Qualifiers
/oranism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/tissue_type="immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imblbing"
/lab_host="E. coli XL0LR"
/note="Organ: juvenile vegetative shoots; Vector:
PAD-GAL4-2.1; Site1: EcorI; Site2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled, polyA+ RNA isolated, and cDNA synthesized for
EcorI (5') and XhoI (3') directional cloning into Lambda
Hybridz vector from Stratagene. Tissue Sources: 1. Whole
shoots 3 days after sowing/imblbing in wet soil. 2. Basal
1.5 cm shoots 6 days after sowing - includes yellow
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex, including adult leaf
primordia, 9 days after sowing. 4. Partially expanded and
greening leaves 4-5 at 13 days after sowing."
BASE COUNT 111 a 108 c 112 g 119 t
ORIGIN
alignment_scores:
Quality: 75.50 Length: 94
Ratio: 1.606 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 25.532
alignment_block:
US-10-048-197-2 x B1698181 ..

```


Align seg 1/1 to: BI698181 from: 1 to: 450

[illegible]

seq_documentation_block:

LOCUS		459 bp	mRNA	linear	EST 13-SEP-1999
DEFINITION	6140681075	- root	cdna	library from walbot lab zea mays	cdna,
DESCRIPTION	mRNA sequence.				

ACCESSION	AW018175	
VERSION	AW018175.1	GI:5871704
REVISION	1	

SOURCE Zea mays.

ORGANIS

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.

REFERENCE	1 (bases 1 to 459)
AUTHORS	Walbot,V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL Unpublished (1999)

COMMENT
Contact: Walbot V

Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 3337

Email: walbot@stanford.edu
Plate: 614068 row: C column: 08.

FEATURES	Location/Qualifiers
source	1. .459

```

/organism="Hsa_mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="614 - root cDNA library from Walbot Lab"
/tissue_type="root"
/dev_stage="3-4 days old"
/lab_host="XLOLR"
/notes="Organ: root; Vector: pBluescriptII SK+; Site:1;
ECORI; Site_2: XhoI; 3-4 days old root tissue from Walbot
Lab (LM)"

```

```

alignment_scores:
  Quality: 75.50      Length: 94
  Ratio: 1.606      Gaps: 4
  Percent Similarity: 50.000      Percent Identity: 25.532

```

```
alignment_block:
```

US-10-048-197-2 x AW018175/rev ..

Align seg. 1/1 to reverse of: AW018175 from: 1 to: 459

```

7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuValThrGlyCysVal
::: |||::: ||| ||| ||
259 AAGGCACCGCTCAAGTCCAGCTCAAGAAAGGCTTTGCATGGCC...GT 213
23 LserThrGlyAsnValAlaMetLysGlu...GlnAsnGlnThrIleG 39
::: |||::: ||| ||| |||::: |||
212 CGCCGTCGCAACTTGGCGATGGAGGAAGACGATCCGCGCAACATCC 163
39 LucGlnThrIleIleLysGlyLysThrAsnLysGlnGlnLysSerArg 55
::: |||::: |||
162 AGATG..... 158
56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG 72
|||::: |||::: |||::: |||
157 .....ACCGTCAACATCTCTTGTTCCCTCTCAAG..... 128
72 yHisThrAlaIleLeuValaProAsnArgTropGlnIleLeuSerLeu 89
::: |||::: |||::: |||::: |||
127 .....AAGAACTGGCGAATGTGAGATGCCCTCT 100

```

seq_name: gb_est1:AI834558

LOCUS	A1834558	460 bp	mrna	linear	EST 02-FEB-2000
DEFINITION	606069H03.x1 606		library	from Schmidt	lab zea
	mays cDNA, mRNA sequence.				

ACCESSION	AI834558	
VERSION	AI834558.1	GI:5468767

KEYWORDS	EST.
SOURCE	Zea mays.

ORGANIS

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 460)

AUTHORS	TITLE
Walbot, V.	Maize ESTs from various cDNA libraries sequenced at Stanford

University
Unpublished (1999)

COMMENT
Contact: walbot v

Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227

Email: walbot@stanford.edu
Plate: 606069 row: H column: 03.

FEATURES	Location/Qualifiers
source	1. .460

```

/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_idb="606 - Ear tissue cDNA library from Schmidt lab"
/tissue_type="mixed"
/dev_stage="ear length from 0.5 cm - 2.0 cm"
/lab_host="XtOLR (Stratagene)"
/notes="Organ: Immature ear; Vector: pBK-CMV; Site: 1: EcorRI; Site: 2: XhoI; Mixed ear tissue cDNA library from Schmidt lab"

```

BASE COUNT	119 a	112 c	97 g	132 t
ORIGIN				

alignment_scores:

Quality: 75.50 Length: 94
Ratio: 1.606 Gaps: 4
Percent similarity: 50.000 Percent identity: 25.532

alignment_block:

US-10-048-197-2 x A1834558/rev ..

Align seg 1/1 to reverse of: A1834558 from: 1 to: 460

```

7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCysVa 23
::: |||::: |||::: |||::: |||::: |||
397 AAGGACACCGTCAGTTCACGCTCAAGAGCTCTTCGATGGC...GT 351

23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
::: |||::: |||::: |||::: |||::: |||
350 CGCCCTCGGCACTTGGCGATGAGAGACGATCCAGCAACATCC 301

39 lGlnThrIleLeuLysGlyThrAsnLysGlnGlnIleSerSerArg 55
::: |||::: |||::: |||::: |||::: |||
300 AGATG..... 296

56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG1 72
|||::: |||::: |||::: |||::: |||
295 .....AGCGTAACCTCCTGTTCCTCCTGCAAG..... 266

72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGlnIleLeuSerLeu1 89
::: |||::: |||::: |||::: |||::: |||
265 .....AAGAACTGCGCAGATGCGATGCGATGCTTT 238

89 lIleSerPheLeuTrpValLysProTyrArg 99
::: |||::: |||::: |||::: |||::: |||
237 ACGTCACAGACACCATGGGGAAGCCATACAGG 206

```

seq_name: gb_est2:BI245316

seq_documentation_block:

LOCUS BI245316 467 bp mRNA linear EST 13-JUN-2001
DEFINITION 949031A01.x2 949 - Juvenile leaf and shoot cDNA from Steve Moose

ACCESSION BI245316
VERSION BI245316.1 GI:14717666
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.

REFERENCE 1. (bases 1 to 467)
AUTHORS Walbot V
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949031 row: A column: 01.

FEATURES

Source location/Qualifiers
1..467
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve Moose"
/tissue_type="Immature leaf primordium and vegetative meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL0LR"

BASE COUNT

116 a 111 c 116 g 124 t

alignment_scores:

Quality: 75.50 Length: 94
Ratio: 1.606 Gaps: 4
Percent similarity: 50.000 Percent identity: 25.532

alignment_block:

US-10-048-197-2 x BI245316 ..

Align seg 1/1 to: BI245316 from: 1 to: 467

```

7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCysVa 23
::: |||::: |||::: |||::: |||::: |||
129 AAGGACACCGTCAGTTCACGCTCAAGAGCTCTTCGATGGC...GT 175

23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
::: |||::: |||::: |||::: |||::: |||
176 CGCCCTCGGCACTTGGCGATGAGAGACGATCCAGCAACATCC 225

39 lGlnThrIleLeuLysGlyThrAsnLysGlnGlnIleSerSerArg 55
::: |||::: |||::: |||::: |||::: |||
226 AGATG..... 230

56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG1 72
|||::: |||::: |||::: |||::: |||
231 .....AGCGTAACCTCCTGTTCCTCCTGCAAG..... 260

72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGlnIleLeuSerLeu1 89
::: |||::: |||::: |||::: |||::: |||
261 .....AAGAACTGCGCAGATGCGATGCGATGCTTT 288

89 lIleSerPheLeuTrpValLysProTyrArg 99
::: |||::: |||::: |||::: |||::: |||
289 ACGTCACAGACACCATGGGGAAGCCATACAGG 320

```

seq_name: gb_est2:BE511166

seq_documentation_block:

LOCUS BE511166 488 bp mRNA linear EST 07-AUG-2000
DEFINITION 946058D02.x1 946 - tassal primordium prepared by schmidt lab zea

ACCESSION BE511166
VERSION BE511166.1 GI:9732414
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.

REFERENCE 1. (bases 1 to 488)
AUTHORS Walbot V
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University

/note="Organ: juvenile vegetative shoots; Vector: PAD-GAL4-2.1; Site.1: EcorI; Site.2: XhoI; Equal amounts of total RNA by weight from 4 tissue sources (see below) were pooled, polyA+ RNA isolated, and cDNA synthesized for EcorI (5') and XhoI (3') directional cloning into lambda HybriZap vector from Stratagene. Tissue Sources: 1. Whole shoots 3 days after sowing/imbibing in wet soil. 2. Basal 1.5 cm shoots 6 days after sowing - includes yellow portions of developing leaves 1-5, primordia from 6-8, and the vegetative apex. 3. Non-green portions of developing leaves 4-5 and the vegetative apex, including adult leaf primordia, 9 days after sowing. 4. Partially expanded and greening leaves 4-5 at 13 days after sowing."

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.

```

71 .....AAGACTGGCAGATGTGAGATGCCTCT 14
72 yhisthralaleuaplaProasnaargtTpglngluileuSerleu 89
      :::::|||||:::  ||

```

```

89 leIleSerPheLeuTrpValIysProTyrArg 99
      ::::: |||||

```

143 ACGTCAAGACCAACCATGGGAGGACCATACAGG 112
 seq_name: gb_est1:AM037170

seq_documentation_block:
 LOCUS AM037170 492 bp mRNA linear EST 15-SEP-1999
 DEFINITION 614018H02.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
 mRNA sequence.
 ACCESSION AM037170
 VERSION AM037170.1 GI:5895924
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 492)
 AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 UNIVERSITY
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 614018 row: H column: 02.

FEATURES
 source
 1. 492
 /organism="Zea mays"
 /cultivar="W23"
 /db_xref="taxon:4577"
 /clone_lib="614 - root cDNA library from Walbot Lab"
 /tissue_type="root"
 /dev_stage="3-4 days old"
 /lab_host="XLOLR"
 /note="Organ: root; Vector: pBluescriptII SK+; Site:1;
 EcoRI; Site:2: XhoI; 3-4 days old root tissue from Walbot
 Lab (LM)"
 BASE COUNT 124 a 116 c 125 g 126 t 1 others
 ORIGIN

alignment_scores:
 Quality: 75.50 Length: 94
 Ratio: 1.606 Gaps: 4
 Percent Similarity: 50.000 Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AM037170 ..

Align seg 1/1 to: AM037170 from: 1 to: 492

```

7 ArgGleuThrIleSerAlaLeuLeuThrAlaLeuValThrlGlyCysVa 23
   ::  |||:::  ||  ::|||  |||  ||
154 AAGGCAACCGTCAAGTTCACAGTCACAGAGGCTTGCATGGC...GT 200
   ::  |||:::  ||  ::|||  |||  ||
23 lserThrGlyAsnValAlaIleMetLysGlu...GlnAsnGlnGlnThrlIleG 39
   ||:::  |||:::  |||  |||  |||  |||  |||  |||  |||  |||
201 CCGCGCGGCGCAACTTGCGATGGAGAGACGATCCACAGCAACATCC 250
   ::  |||:::  ||  ::|||  |||  |||  |||  |||  |||  |||  |||
39 lGlnThrIleIleLysGlyLysThrAsnLysGlnGlnLysSerSerArg 55
   ::  |||:::  ||  ::|||  |||  |||  |||  |||  |||  |||  |||
251 ACATG..... 255
56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG 72
   ||:::  |||:::  |||  |||  |||  |||  |||  |||  |||  |||
256 .....AGCGTGAACCTCTGTTCCCTCTGTAG..... 285
72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGlnIleLeuSerLeu 89
   ::  |||:::  ||  ::|||  |||  |||  |||  |||  |||  |||  |||

```

286AAGAACTGGCAGAAATGTGATGCCTCT 313
 89 leIleSerPheLeuThrIleValIleProTyrArg 99
 :::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 314 ACGTCAAGACCAACCATGGGAGGACCATACAGG 345
 seq_name: gb_est2:B1478440

seq_documentation_block:
 LOCUS B1478440 512 bp mRNA linear EST 27-AUG-2001
 DEFINITION 949069H08.y1 949 - juvenile leaf and shoot cDNA from Steve Moose
 Zea mays cDNA, mRNA sequence.
 ACCESSION B1478440
 VERSION B1478440.1 GI:15312868
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 512)
 AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 UNIVERSITY
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 949069 row: B column: 08.

FEATURES
 source
 1. 512
 /organism="Zea mays"
 /cultivar="W64A"
 /db_xref="taxon:4577"
 /clone_lib="949 - juvenile leaf and shoot cDNA from Steve
 Moose"
 /tissue_type="immature leaf primordium and vegetative
 meristem"
 /dev_stage="4 stages from 3-13 days after imbibing"
 /lab_host="E. coli XLOLR"
 /note="Organ: juvenile vegetative shoots; Vector:
 pAD-GAL4-2.1; Site:1: EcoRI; Site:2: XhoI; Equal amounts
 of total RNA by weight from 4 tissue sources (see below)
 were pooled, polyA+ RNA isolated, and cDNA synthesized for
 EcoRI (5') and XhoI (3') directional cloning into lambda
 HybriZap vector from Stratagene. Tissue Sources: 1. Whole
 shoots 3 days after sowing/imbibing in wet soil. 2. Basal
 1.5 cm shoots 6 days after sowing - includes yellow
 portions of developing leaves 1-5; primordia from 6-8; and
 the vegetative apex. 3. Non-green portions of developing
 leaves 4-5 and the vegetative apex, including adult leaf
 primordia, 9 days after sowing. 4. Partially expanded and
 greening leaves 4-5 at 13 days after sowing."

BASE COUNT 139 a 122 c 126 g 125 t
 ORIGIN

alignment_scores:
 Quality: 75.50 Length: 94
 Ratio: 1.606 Gaps: 4
 Percent Similarity: 50.000 Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x B1478440 ..

Align seg 1/1 to: B1478440 from: 1 to: 512

```

7 ArgGleuThrIleSerAlaLeuLeuThrAlaLeuValThrlGlyCysVa 23
   ::  |||:::  ||  ::|||  |||  |||  |||  |||  |||  |||  |||

```

```

148 AAGGCAGACGGTCGCAAAATTCAGCTCAAGAGAGGTCCTTGATGGC...GT 194
23 IserthrglyasnValAlaMetLysGlu...GlnasngInglThrIleG 39
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
195 CCCGCGCGCAACTTGTCGATGAGAGAGCAAGCATCCAGCAAGACATCC 244
39 InGIThrIleIleLysGlyLysThrAsnLysGlnIleLysSerArg 55
||||:
245 AGATG..... 249
56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG 72
||||:||||:||||:||||:||||:||||:||||:||||:||||:
250 .....AGCGTGAACCTCCCTGTGTTCCCTCTGAG..... 279
72 yHsThrAlaIleLeuAlaProAsnArgTrpGlnIleIleSerLeuI 89
||||:||||:||||:||||:||||:||||:||||:||||:||||:
280 .....AAGACATGCGCAGATGTGATGATGCTCT 307
89 IeIleSerPheLeuTrpValLysProTyrArg 99
||||:||||:||||:||||:||||:||||:||||:||||:||||:
308 ACGTCAGAGACACCATGGGAAAGCCATACAGC 339

seq_name: gb_est1:AW171836

seq_documentation_block:
LOCUS AW171836 513 bp mRNA linear EST 12-NOV-1999
DEFINITION 618051811.x1 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA
sequence.
ACCESSION AW171836
VERSION AW171836
KEYWORDS AW171836.1 GI:6403361
EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade: Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 513)
REFERENCE
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 618051 row: B column: 11.
FEATURES
Source
1..513
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="618 - Inbred Tassel cDNA Library"
/tissue_type="tassel"
/dev_stage="tassel"
/lab_host="X10LR"
/length=length from 0.1 to 2.5 cm"
/notes="Organ: tassel; Vector: PAD-GAL4-2.1 (Hybridzap);
Inbred tassel library from Schmidt lab"
BASE COUNT 114 a 137 c 128 g 134 t
ORIGIN
Alignment_scores:
Quality: 75.50 Length: 94
Ratio: 1.606 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 25.532

alignment_block:
US-10-048-197-2 x AW171836/rev ..
Align seg 1/1 to reverse of: AW171836 from: 1 to: 513

```

```

7  ATgletThrlleSerAlaLeuThrAlaLeuValThrclycys 23
   ::: |||:::  |||  ::||  |||  ||
240  AAGGCAACCGTCAGTCCAGCTCAAGAAGGTCCTTGCAATGGGC...GT 194
23  lserThrclyAsnValAlaMetLysGlu...GlnAengInGlnTrileo 39
   |||:::|||||:::|||||  ||  |||  |||:::|||||
193  CGCCGTCGGCAACTTCGCATGGAGAGAAGCAGATCCAGCAGACATCC 144
39  luginThrIleIleLysGlyLysThrAsnLysInLuleSerArg 55
   ::::
143  AGATG..... 139
56  PhcGlySerAlaAspSerIleSerPheMetIleValIleLysPheG 72
   |||:::|||||:::|||||  ::|||
138  .....ACGCTGAACCTCCTGTTTCCCTCTGAG..... 109
72  yHstHrAlaIleLeuAlaProAsnArGrTgInGluIleLeuSerLeu 89
108  .....AAGACACGGCAGAAATGTGATGCTCT 81
89  lelleSerPheLeuTrpValLysProTyrArg 99
   ::::  |||  |||  |||  |||
80  ACGTCAAGACGACCATGGGAAACCATACAGC 49

seq_name: gb_est1:AI586595

seq_documentation_block:
LOCUS      AI586595                      514 bp    mRNA    linear    EST 07-Apr-1999
DEFINITION  486049C06.x4 486 - leaf primordia cDNA library from Hake lab Zea
            mayS cDNA, mRNA sequence.
ACCESSION  AI586595
VERSION    AI586595.1  GI:4572946
KEYWORDS   EST.
SOURCE     Zea mays.
            Zea mays.
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
1         1 (bases 1 to 514)
REFERENCE  1  Walbot,V.
            Make ESTs from various cDNA libraries sequenced at Stanford
            University
AUTHORS    Unpublished (1999)
COMMENT    Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 8227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 486049 row: C column: 06.
FEATURES
    source
        1..514
            /organism="Zea mays"
            /cultivar="B73"
            /db_xref="taxon:4577"
            /clone_lib="486 - leaf primordia cDNA library from Hake
            lab"
            /tissue_type="leaf primordia"
            /dev_stage="p7-p11 leaf"
            /lab_host="E.coli XL1-Blue MFR"
            /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
            library"
BASE COUNT      138 a      123 c      113 g      140 t
ORIGIN
alignment_scores:
    Quality:      75.50      Length:      94
    Ratio:        1.606      Gaps:        4
    Percent Similarity: 50.000      Percent Identity: 25.532

```

alignment_block:
US-10-048-197-2 x A1586595/rev ..

Align seg 1/1 to reverse of: A1586595 from: 1 to: 514

```

7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCysVa 23
  :: |||::: ||| ::|||
376 AAGCAACCGTCAGATTCACAGAGAGGCTTCATGAGGT...GT 330
23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIle 39
  |:::|||||:::|||||:::|||||
329 TGGCGTTGGCACTTGGCGATGGAGAGAGATCCAGACAGAACATCC 280
39 lGlnThrIleIleLysGlyLysThrAsnLysGlnGlnIleSerSerArg 55
  ::::
279 AGATG..... 275
56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG1 72
  |||:::|||||:::
274 ..ACGTCACACTTCCTGTTCCCTTCGAG..... 245
72 yHISThrAlaIleLeuAlaProAsnArgTyrGlnGlnIleLeuSerIleu 89
  ::::|||:::
244 .....AAGAACTGGCAGAAATGTGATGATGCTCT 217
89 lIleSerPheLeuTyrValLysProTyrArg 99
  ::::
216 ACGTCAGAGCACCATGGGGAGGCCATACAG 185

```

seq_name: gb_est2:BI643532

seq_documentation_block:

LOCUS BI643532 518 bp mRNA linear EST 10-SEP-2001
DEFINITION 949079C10.x1 949 - Juvenile leaf and shoot cDNA from Steve Moose

ACCESSION BI643532
VERSION BI643532.1 GI:15545738

KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 518)
Walbot, V.

Maize ESTs from various cDNA libraries sequenced at Stanford
University

JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949079 row: C column: 10.
Location/Qualifiers
1. 518

FEATURES
source
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_id="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/tissue_type="Immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL0LR"
/note="Organ: juvenile vegetative shoots; Vector:
pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled, polyA+ RNA isolated, and cDNA synthesized for
EcoRI (5') and XhoI (3') directional cloning into lambda
Hybridap vector from Stratagene. Tissue Sources: 1. Whole

shoots 3 days after sowing/imbibing in wet soil. 2. Basal
1.5 cm shoots 6 days after sowing - includes yellow
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex. Including adult leaf
primordia, 9 days after sowing. 4. Partially expanded and
greening leaves 4-5 at 13 days after sowing."

BASE COUNT 138 a 130 c 122 g 128 t

alignment_scores:
Quality: 75.50 Length: 94
Ratio: 1.606 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 25.532

alignment_block:
US-10-048-197-2 x BI643532/rev ..

Align seg 1/1 to reverse of: BI643532 from: 1 to: 518

```

7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCysVa 23
  :: |||::: ||| ::|||
358 AAGCAACCGTCAGATTCACAGAGGCTTCATGAGGC...GT 312
23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIle 39
  |:::|||||:::|||||:::|||||
311 CGCGTGGCACTTGGCGATGGAGAGAGACAGATCCAGACAGAACATCC 262
39 lGlnThrIleIleLysGlyLysThrAsnLysGlnGlnIleSerSerArg 55
  ::::
261 AGATG..... 257
56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG1 72
  |||:::|||||:::
256 ..ACGTCACACTTCCTGTTCCCTTCGAG..... 227
72 yHISThrAlaIleLeuAlaProAsnArgTyrGlnGlnIleLeuSerIleu 89
  ::::|||:::
226 .....AAGAACTGGCAGAAATGTGATGATGCTCT 199
89 lIleSerPheLeuTyrValLysProTyrArg 99
  ::::
198 ACGTCAGAGCACCATGGGGAGGCCATACAG 167

```

seq_name: gb_est1:A1629856

seq_documentation_block:

LOCUS A1629856 529 bp mRNA linear EST 26-APR-1999
DEFINITION 486041B10.x2 486 - leaf primordia cDNA library from Hake lab Zea
mays cDNA, mRNA sequence.

ACCESSION A1629856
VERSION A1629856.1 GI:4681186

KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 529)
Walbot, V.

Maize ESTs from various cDNA libraries sequenced at Stanford
University

JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 486041 row: B column: 10.
Location/Qualifiers

FEATURES

```
source 1..529
        /organism="Zea mays"
        /cultivar="B73"
        /db_xref="taxon:4577"
        /clone_idb="486" - leaf primordia cDNA library from Hake
        lab"
        /tissue_type="leaf primordia"
        /dev_stage="p7-p11 leaf"
        /lab_host="E.coli XL1-Blue MFR"
        /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
        library."

BASE COUNT 142 a 126 c 114 g 147 t
ORIGIN

alignment_scores:
    Quality: 75.50      Length: 94
    Ratio: 1.606        Gaps: 4
    Percent Similarity: 50.000    Percent Identity: 25.532

alignment_block:
US-10-048-197-2 x A1629856/rev ..

Align seg 1/1 to reverse of: A1629856 from: 1 to: 529

7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
   :::::::::::::: :::: ::::: ::::: :::::
397 AAGGCAACCCCTCAAGTTCAGCTCAGAGAGAGTCTTTCATGGGT...GT 351

23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
   :::::::::::::: ::::: ::::: ::::: :::::
350 TGGCGTTGGCAACTTGGCGATGAGAGAGAGATCCACAGACATCC 301

39 lGlnThrIleIleLysGlyLysThrAsnLysGlnGlnIleSerSerArg 55
   :::::
300 AGATG..... 296

56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheG1 72
   :::::::::::::: ::::: ::::: :::::
295 .....AGCGTCAACTTCCTGTTGCTTCCCTCTGAAG.... 266

72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGlnIleLeuSerLeuI 89
   :::::::::::::: ::::: ::::: :::::
265 .....AAGAACTGGCAAGATGTGATGATGCTCT 238

89 lIleSerPheLeuTrpValLysProTyrArg 99
   :::::
237 ACGTCAAGAGCACCATGGGAGGAGCATACAGG 206

seq_name: gb_est2:B1180391

seq_documentation_block:
LOCUS B1180391 533 bp mRNA linear EST 09-JUL-2001
DEFINITION 949031A01.x1 949 - Juvenile leaf and shoot cDNA from Steve Moose
Zea mays cDNA, mRNA sequence.
ACCESSION B1180391
VERSION B1180391.1 GI:14646202
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 533)
Walbot, V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
```

```
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949031 row: A Column: 01.
Location/Qualifiers
1..533
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_idb="949" - Juvenile leaf and shoot cDNA from Steve
Moose"
/tissue_type="Immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL10R"
/note="Organ: juvenile vegetative shoots; Vector:
PAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled, polyA+ RNA isolated, and cDNA synthesized for
EcoRI (5') and XhoI (3') directional cloning into lambda
Hybridzap vector from Stratagene. Tissue Sources: 1. Whole
shoots 3 days after sowing/imbibing in wet soil. 2. Basal
1.5 cm shoots 6 days after sowing - includes yellow
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex, including adult leaf
primordia, 9 days after sowing. 4. Partially expanded and
greening leaves 4-5 at 13 days after sowing."

BASE COUNT 142 a 133 c 124 g 134 t
ORIGIN

alignment_scores:
    Quality: 75.50      Length: 94
    Ratio: 1.606        Gaps: 4
    Percent Similarity: 50.000    Percent Identity: 25.532

alignment_block:
US-10-048-197-2 x B1180391/rev ..

Align seg 1/1 to reverse of: B1180391 from: 1 to: 533

7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
   :::::::::::::: ::::: ::::: :::::
364 AAGGCAACCCCTCAATTCAGCTCAGAGAGAGTCTTTCATGGGC...GT 318

23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
   :::::::::::::: ::::: ::::: :::::
317 CGCGTGGCACTTGGCGATGAGAGAGAGATCCACAGACATCC 268

39 lGlnThrIleIleLysGlyLysThrAsnLysGlnGlnIleSerSerArg 55
   :::::
267 AGATG..... 263

56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheG1 72
   :::::::::::::: ::::: ::::: :::::
262 .....AGCGTCAACTTCCTGTTGCTTCCCTCTGAAG.... 233

72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGlnIleLeuSerLeuI 89
   :::::::::::::: ::::: ::::: :::::
232 .....AAGAACTGGCAAGATGTGATGCTCT 205

89 lIleSerPheLeuTrpValLysProTyrArg 99
   :::::
204 ACGTCAAGAGCACCATGGGAGGAGCATACAGG 173

seq_name: gb_est1:AM090901

seq_documentation_block:
LOCUS AM090901 554 bp mRNA linear EST 18-OCT-1999
DEFINITION 614068C08.y1 614 - root cDNA library from Walbot lab Zea mays cDNA,
mRNA sequence.
ACCESSION AM090901
VERSION AM090901.1 GI:6056511
```

KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 554)
AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY University of California, Stanford
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 723 8221
Email: walbot@stanford.edu
Plate: 614068 row: C column: 08.
Location/Qualifiers

FEATURES
source 1..554
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="614 - root cDNA library from Walbot Lab"
/tissue_type="root"
/dev_stage="3-4 days old"
/lab_host="XLOLR"
/note="Organ: root; Vector: pBluescriptII SK+; Site_1:
ECORI; Site_2: XhoI; 3-4 days old root tissue from Walbot
Lab (LM)"
BASE COUNT 143 a 129 c 142 g 140 t
ORIGIN

alignment_scores:
Quality: 75.50 Length: 94
Ratio: 1.606 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 25.532

alignment_block:
US-10-048-197-2 x AM090901 ..
Align seg 1/1 to: AM090901 from: 1 to: 554

```

7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCysVa 23
::: |||::: |||
218 AAGGCAACCGCAATTCACGCTCAAGAGAGTCCTTGCATGGGC...GT 264
23 lserThrgIyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
|:::|||||:::|||||::: ||| ||| ||| ||| |||
265 CGCCGTCGGCAACTTGGCGATGAGAGAGACGATCCGCGAAGCATCC 314
39 lueGlnThrIleIleLysGlyLysThrAsnLysGlnGlnLysSerArg 55
::: |||::: |||
315 AGATG..... 319
56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG 72
|||:::|||||::: |||
320 .....AGCGTGAACCTTCTTCCCTTCGTAAG..... 349
72 yHsthrAlaIleLeuAlaProAsnArgTrpGlnGlnIleLeuSerIle 89
::: |||::: |||
350 .....AAGAACTGGCAATGTGAGATGAGATCGCTCT 377
89 lseIleSerPheLeuThrValLysProTyrArg 99
::: |||::: |||
378 ACGTCAAGACGACCATGGGAAGCATACAGG 409
seq_name: gb_est2:BI675165
seq_documentation_block:
LOCUS BI675165 557 bp mRNA linear EST 12-SEP-2001

```

DEFINITION 949079C10.x2 949 - Juvenile leaf and shoot cDNA from Steve Moose
Zea mays cDNA, mRNA sequence.
ACCESSION BI675165
VERSION BI675165.1 GI:15590549
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 557)
AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY University of California, Stanford
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 723 8221
Email: walbot@stanford.edu
Plate: 949079 row: C column: 10.
Location/Qualifiers

FEATURES
source 1..557
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/tissue_type="Immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XLOLR"
/note="Organ: juvenile vegetative shoots; Vector:
pAD-GAL4-2.1; Site_1: EcorI; Site_2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled, polyA+ RNA isolated, and cDNA synthesized for
EcorI (5') and XhoI (3') directional cloning into lambda
Hybridap vector from Stratagene. Tissue Sources: 1. Whole
shoots 3 days after sowing/imbibing in wet soil. 2. Basal
1.5 cm shoots 6 days after sowing - includes yellow
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex, including adult leaf
primordia. 9 days after sowing. 4. partially expanded and
greening leaves 4-5 at 13 days after sowing."

BASE COUNT 143 a 143 c 124 g 147 t
ORIGIN

alignment_scores:
Quality: 75.50 Length: 94
Ratio: 1.606 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 25.532

alignment_block:
US-10-048-197-2 x BI675165/rev ..
Align seg 1/1 to reverse of: BI675165 from: 1 to: 557

```

7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCysVa 23
::: |||::: |||
337 AAGGCAACCGCAATTCACGCTCAAGAGAGTCCTTGCATGGGC...GT 291
23 lserThrgIyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
|:::|||||:::|||||::: ||| ||| ||| ||| |||
290 CGCCGTCGGCAACTTGGCGATGAGAGAGACGATCCAGCAACATCC 241
39 lueGlnThrIleIleLysGlyLysThrAsnLysGlnGlnLysSerArg 55
::: |||::: |||
240 AGATG..... 236

```



```

56 PhegylSerAlaspserIleSerPheNetIValIleLysPhecl 72
235 .....AGCGTGAACCTCCTGTGTTCCCTGTAAG..... 206
|||||
72 yHlStrAlalAleAAlProAsnArGrPrpsIngluileLeuSerLeui 89
205 .....AGAACTCGGAGAATCGAATCGCTCT 178
|||||
seq_name: gb_eest1:A1714378

seq_documentation_block:
LOCUS A1714378 559 bp mRNA linear EST 02-FEB-2000
DEFINITION 606013A04.x2 606 - Ear tissue cDNA library from Schmidt lab zea
mays CDNA, mRNA sequence.
VERSION A1714378
KEYWORDS A1714378.1 GI:5018185
SOURCE EST.
ORGANISM Zea mays.
Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1..to 559)
Maibot.V.
REFERENCE Maize ESTs from various cDNA libraries sequenced at Stanford
AUTHORS University
TITLE Unpublished (1999)
JOURNAL Contact: Walbot V
COMMENT Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 606013 row: A column: 04.

FEATURES
source location/Qualifiers
1..559
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/cclone_lib="606 - Ear tissue cDNA library from Schmidt
lab"
/tissue_type="mixed"
/dev_stage="ear length from 0.5 cm - 2.0 cm"
/lab_host="XLOLR (Stratagene)"
/note="Organ: Immature ear; Vector: pBK-CMV; Site_1: EcoRI
; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
lab"

BASE COUNT 144 a 136 c 126 g 153 t
ORIGIN
Alignment_scores:
Quality: 75.50 Length: 94
Ratio: 1.606 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 25.532

alignment_block:
US-10-048-197-2 x A1714378/rev ..
Align seg 1/1 to reverse of: A1714378 from: 1 to: 559
7 ArgLeuThrIleSerAlaleuLeuThralaleuLeuValThrGlycysVa 23
::: ||||| ::||| |||
393 AAGGCACCGTCAGATTCCAGCTCAAGAGAGTCCTTGATGGGC...GT 347
23 lserThrGlyAsnValAlAlamellysgLu...GlnasngIngnThrllleg 39
|||||
346 CGCGTCGGCAACTTGGCGATGGAGGAGAGCATGCCAGCAACAATC 297

```

```

39 InG1HrThrlleIeLysGlyLysTThrAsnLysInclInLysSerArg 55
296 AGTGTG..... 292
56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG1 72
291 .....AGCGTGAACCTTCCTGTTCCCTCTGTAAG..... 262
72 yH1stHrAlaIleLeuAlaProbsnArGrtpGInG1uIleLeuSerLeu 89
261 .....AAGAACTGGCGAGAACTGTGATGCATCCCTC 234
89 leIleSerPheLeuTrpValLysProTyrArg 99
233 ACCTCAAGAGACACCATGGGAGGCCATPACAGG 202
seq_name: gb_est1:AW566286
seq_documentation_block:
LOCUS AW566286 563 bp mRNA linear EST 10-MAR-2000
DEFINITION 660069A03.y1 660 - Mixed stages of anther and pollen Zea mays cDNA.
ACCESSION AM566286
VERSION AM566286
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
REFERENCE 1 (bases 1 to 563)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660069 row: A column: 03.
FEATURES
Source location/Qualifiers
1..563
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="660 - Mixed stages of anther and pollen"
/tissue_type="whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="X10LR"
/note="Organ: anthers; Vector: Lambda Zap; Site_1: EcorI; Site_2: XhoI; Anther and pollen cDNA library. Directionally sequenced with 5' end at the EcorI site. Created by Amie Franklin."
BASE COUNT 146 a 136 c 153 g 128 t
ORIGIN
alignment_scores:
Quality: 75.50 Length: 94
Ratio: 1.606 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 25.532
alignment_block:
US-10-048-197-2 x AW566286 ..
Align seg 1/1 to: AW566286 from: 1 to: 563
7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrIlycysVa 23
::: |||::: ||| ::||| |||

```

```

39  InG1ThrlleIeLysGlyStHraSnLysInclInLieserrArg 55
      :::::
296  AGTGTG.....
56  pNeglySerAlaAspSerIleSerPheMetIleValIleLysPheG1 72
      |||:::|||||:::|:::|:::|:::|:::|:::|:::|:::|
291  .....AGCGTGAACCTTCCTGTTCCCTCTGTAAG..... 262
72  yHstHraIAlaIleuAlaProbsnArGrtpngInG1uIleLeuSerLeu 89
      :::::|||||:::|:::|:::|:::|:::|:::|:::|:::|
261  .....AAGAACGTGGCAGAACTGTGATGCATCCCTCT 234
89  leIleSerPheLeuTrpValLysProTyrArg 99
      :::::|||||:::|:::|:::|:::|:::|:::|:::|:::|
233  ACCTCAAGAGACACCATGGGAGGCCATPACAGG 202

seq_name: gb_est1:AW566286

seq_documentation_block:
LOCUS      AW566286                    563 bp      mRNA      linear      EST 10-MAR-2000
DEFINITION  660069A03.y1 660 - Mixed stages of anther and pollen Zea mays cDNA.
ACCESSION  AM566286
VERSION    AM566286
KEYWORDS   Zea mays.
SOURCE     Zea mays.
ORGANISM   Zea mays.
REFERENCE  1. (bases 1 to 563)
AUTHORS    Walbot,V.
TITLE       Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL    Unpublished (1999)
COMMENT    Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 660069 row: A column: 03.

FEATURES             Location/Qualifiers
     source            1..563
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="660 - Mixed stages of anther and pollen"
                     /tissue_type="whole premeiotic anthers to pollen shed"
                     /dev_stage="premeiotic anthers to pollen shed"
                     /lab_host="X10LR"
                     /note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
                     Site_2: XhoI; Anther and pollen cDNA library.
                     Directionally sequenced with 5' end at the EcoRI site.
                     Created by Amie Franklin."
BASE COUNT      146 a      136 c      153 g      128 t
ORIGIN
alignment_scores:
      Quality:      75.50      Length:      94
      Ratio:        1.606      Gaps:      4
      Percent Similarity: 50.000      Percent Identity: 25.532

alignment_block:
US-10-048-197-2 x AW566286 ..
Align seg 1/1 to: AW566286 from: 1 to: 563
::: |||:::|:::|:::|:::|:::|:::|:::|:::|
7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrIlycysVa 23

```


Tel: 515-294-0975
Fax: 515-294-2299
Email: schnable@state.edu

Individual basecall and confidence value were assigned using the phred software,
(<http://depts.washington.edu/ventures/collabtr/direct/index.htm#b>
rt). Overall sequence quality assessment and vector trimming were
conducted using the Lucy software (<http://www.tigr.org/softlab/lucy/>).
Lucy parameters were set to ensure an overall trimmed quality of
97.5% or better without any vector fragments in the chosen
high-quality region of each sequence. Low-quality bases between the
poly-T and the high-quality region were replaced with N's to serve
as spacers.

PCR Primers

FORWARD: Forward PCR primer sequence, primer T7-1 (AA TAC GAC TCA
CTA TAG)
CTA TAG)

BACKWARD: Backward PCR primer sequence, primer T3 (ATT AAC CCT CAC
TAA AG)

Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

FEATURES

source

1..607

/organism="Zea mays"

/cultivar="B73"

/db_xref="taxon:4577"

/clone="MEST96-F07"

/clone_lib="ISUM4-TN"

/tissue_type="Seedling and silk"

/lab_host="DH10B"

/note="Vector: pT7T3PAC; Site_1: EcoRI; Site_2: NotI;
ds-cDNA molecules were generated as follows. First-strand
cDNA was prepared from oligo-dT selected mRNA by priming
with a NotI oligo-dT primer (5',
AAGCGAGAAATTCGGCGCGAGCAATTTTATTTT). The
resulting DNA:RNA hybrid was treated with RNase H and used
as a template for DNA Pol-catalyzed second strand
synthesis. After the addition of EcoRI adaptors, the
ds-cDNAs were digested with NotI and size-selected. The
resulting molecules were directionally cloned into the
EcoRI and NotI sites of the pT7T3PAC vector. The library
then went through one round of normalization to Cot value
of 5 based on the methods of Marcelo Bento Soares (Genome
Research 6: 791-806, 1996)."

BASE COUNT

158 a 144 c 128 g 177 t

ORIGIN

alignment_scores:

Quality: 75.50 Length: 94
Ratio: 1.606 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 25.532

alignment_block:

US-10-048-197-2 x BM079588/rev ..

Align seg 1/1 to reverse of: BM079588 from: 1 to: 607

7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCysVa 23

417 AAGGCAACCGTCAGTCCAGCTCAGACAGGTCTTGCATGGCT...GT 371

23 ISeThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39

370 TGGCGTGGCACTTGGCGATGAGGAGACAGATCAGACAGAACATCC 321

39 LucInThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55

320 AGATG..... 316

56 PhcGlySerAlaAspSerIleSerPheMetIleValIleLysPheG1 72

315AGCGTCAACTTCCTGTGTTCCCTCTGAGG..... 286

72 yHISThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeu 89

285AGGAACTGGCAGAAATGTGATGCCTCT 258

89 leIleSerPheLeuTrpValLysProTyrArg 99

257 ACCTCAAGACACCATGAGGAGCATACAGG 226

seq_name: gb_est1:AW324655

seq_documentation_block:

LOCUS AW324655 611 bp mRNA linear EST 27-JAN-2000

DEFINITION 707036H01.x1 707 - Mixed adult tissues from Walbot lab (SK) Zea

mayes cDNA, mRNA sequence.

ACCESSION AW324655

VERSION AW324655.1 GI:6760556

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoidae; Andropogoneae; Zea.

1 (bases 1 to 611)

Walbot, V.

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 707036 row: H column: 01.

FEATURES

source

1..611

/organism="Zea mays"

/cultivar="W23"

/db_xref="taxon:4577"

/clone_lib="707 - Mixed adult tissues from Walbot lab (SK

)"

/tissue_type="tassel, kernel, silk, husk, leaf"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: tassel, kernel, silk, husk, root, leaf;
Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
differentiated maize tissues from an active Mutator

plant. Tissue ratio is 4/2/1/1/1 (tassel, kernel, silk,
husk, root, leaf). Unidirectionally cloned."

BASE COUNT

155 a 153 c 133 g 170 t

ORIGIN

alignment_scores:

Quality: 75.50 Length: 94
Ratio: 1.606 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 25.532

alignment_block:

US-10-048-197-2 x AW324655/rev ..

Align seg 1/1 to reverse of: AW324655 from: 1 to: 611

7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCysVa 23

388 AAGGCAACGTCAGTCCAGCTCAGAAAGTCTTGCATGGCT...GT 342

23 ISeThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39

341 CGCCGTCGCACTTGGCGATGAGGAGACAGATCAGACAGAACATCC 292

39 LucInThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55

```

291 AGATG..... 287
56 PhcglySerAlaaspSerIleSerPheMetIleValIleIleSpheG1 72
   |||:|||||:|||||:
286 .....AGCGTCAACCTCTGTTGTTCCCTCTGTAAG..... 257
72 yHstHrAlaIleLeuAlaProAsnArgTrpGlnGulIleLeuSerLeuI 89
   |||:|||||:|||||:
256 .....AAGAACGTGGCAGATGTGAGATGCGCTGT 229
89 leIleSerPheLeuTrpValIysProTyrArg 99
   |||:|||||:
228 ACGTCAAGACACCATGGGAAAGCCATACAGC 197

seq_name: gb_est1:A1711637

seq_documentation_block:
LOCUS   A1711637                613 bp    mRNA    linear    EST 02-FEB-2000
DEFINITION  605058EB06.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays/
ACCESSION  A1711637
VERSION    A1711637.1   GI:5005575
SOURCE     EST.
ORGANISM   Zea mays.
            Zea mays.
REFERENCE  Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 613)
AUTHORS    Walbot,V.
TITLE       Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL    Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 605058 row: E column: 06.

FEATURES
            source
            location/Qualifiers
                1..613
                /organism="Zea mays"
                /cultivar="Oh1043"
                /db_xref="taxon:4577"
                /clone.lib="605 - Endosperm cDNA library from Schmidt lab"
                /tissue-type="nucellar, embryo, and endosperm"
                /dev_stage="10-14 days post-pollination"
                /lab_host="DH5(alpha)"
                /note="Organ: kernel; Vector: PAD-GAL4-2'; Site:1: ECORI;
                Site_2: XhoI; kernel endosperm cDNA library from Schmidt
                lab"
BASE COUNT  158 a      158 c      138 g      159 t

ORIGIN
alignment_scores:
    Quality:      75.50      Length:      94
    Ratio:        1.606      Gaps:        4
    Percent Similarity:  50.000      Percent Identity:  25.532

alignment_block:
US-10-048-197-2 x A1711637/rev ..

Align seg 1/1 to reverse of: A1711637 from: 1 to: 613

7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
   :: |||:|||||: ||| :|||: ||| ||
362 AAGCGAACCGTCACAGTTCACGCTCAAGAGAGCCTTTCATGCGGC...GT 316
   |:::|||||:|||||:||||| ||| |||:|||||:|||||:

23 lserThrGlyAsnValAlaMetIleGlu...GlnAsnGlnGlnThrIleG 39
   |:::|||||:|||||:||||| ||| |||:|||||:|||||:

```

```

315 CGCCGTCGGCACTTGGCATGGAGAGAAAGCATCCAGCAGACATCC 266
39 luGlnThrIleIysGlyIysThrAsnLysGlnIuIleSerSerArg 55
   ::::
265 AGATG..... 261
56 PheGlySerAlaAspSerIleSerPheMetIleValIleLysPheCl 72
   |||:::|||||::: 231
260 ..... AGCGTGAACCTCCGTGTCCCTCTGAAAG..... 231
72 yHISThrAlaIleLeuAlaProAsnArgTrpGlnGlnIuIleSerLeuI 89
230 ..... AAGAACGTGGCAATGTGATGATGCCTCT 203
89 leIleSerPheLeuTrpValysProTyrArg 99
   ::::
202 ACGTCAAAGAGCAACCATGGGGAAGCCCATACAGC 171
seq_name: gb_est1:AW927882
seq_documentation_block:
LOCUS AW927882 614 bp mRNA linear EST 30-MAY-2000
DEFINITION 945010D06.Y1.945 - Mixed adult tissues from Walbot lab, same as 707
(SK) Zea mays cDNA, mRNA sequence.
ACCESSION AW927882
VERSION AW927882.1 GI:8103229
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 614)
Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 945010 row: D column: 06.
FEATURES
Source location/Qualifiers
1..614
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="945 - Mixed adult tissues from Walbot lab,
same as 707 (SK)"
/tissue_type="tassel, kernel, silk, husk, root, leaf"
/dev_stage="fully_grown"
/lab_host="DH10B"
/note="Organ: tassel, kernel, silk, husk, root, leaf;
Vector: pGAD10; Site: 1: EcoRI; cDNA library from fully
differentiated maize tissues from an active Mutator plant.
Tissue ratio is 4:2:1:1:1 (tassel, kernel, silk, husk,
root, leaf). Unidirectionally cloned. New library number
given to library 707 for additional sequencing."
BASE COUNT 174 a 133 c 149 g 158 t
ORIGIN
alignment_scores:
Quality: 75.50 Length: 94
Ratio: 1.606 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 25.532
alignment_block:
US-10-048-197-2 x AW927882 ..

```

usually in hybrid. Sample insert size range was 200 bp to 3 kb with a 1 kb average."

```

/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/node="Organ: Kernel; Vector: PAD-GAL4-2'; Site_1: EcoRI;
Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

BASE COUNT      165 a      167 c      140 g      166 t

ORIGIN

alignment_scores:
    Quality:      75.50      Length:      94
    Ratio:        1.606      Gaps:        4
Percent Similarity: 50.000      Percent Identity: 25.532

alignment_block:
US-10-048-197-2 x A1665346/rev ..

Align seg 1/1 to reverse of: A1665346 from: 1 to: 638

7  ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThGlyCysIra 23
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
373  AAGGCAACCGCTCAAGTTCACGCTCAAGAGAGAGCTTTGCATGGC...GT 327

23  IserThrGlyAsnValAlaMetIysGlu...GlnAsnGlnGlnThrIleG 39
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
326  CGCCCTCGGCAACTGGCGATGAGAGAGAGAGATCCAGCAGAACATCC 277

39  IuGlnThrIleIleIysGlyIysThrAsnIysGlnGluIleSerIserArg 55
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
276  AGATG..... 272

56  PheGlySerAlaAspSerIleSerPheMetIleValIleIleIysPheI 72
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
271  .....ACGCTGAACCTCCTGTTCCCTCTGTAAG..... 242

72  yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
   241  .....AAGAGCTGGCAGAGATGTGAGATGCCTCT 214

89  IeIleSerPheLeuTrpValIysProTyrArg 99
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
213  AAGTCAAGACGACCAATGGGGAAGCCATACAGG 182

seq_name: gb_est1:AW313326

seq_documentation_block:
LOCUS      AW313326      650 bp      mRNA      linear      EST 24-JAN-2000
DEFINITION  707024C08.x1.707 - Mixed adult tissues from Walbot lab (SK) Zea
mays cDNA, mRNA sequence.
ACCESSION  AW313326
VERSION    AW313326.1 GI:6742511
KEYWORDS   EST.

SOURCE     Zea mays.
ORGANISM   Zea mays.
Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 650)
Walbot.V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 707024 row: C column: 08.
Location/Qualifiers
1..650
/organism="Zea mays"

```

```

/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="707 - Mixed adult tissues from Walbot lab (SK
)"
/tissue_type="tassel, kernel, silk, husk, root, leaf"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: tassel, kernel, silk, husk, root, leaf;
vector: pGAD10; Site_1: EcoRI; cDNA library from fully
differentiated maize tissues from an active Mutator
plant. Tissue ratio is 4/2/1/1/1 (tassel, kernel, silk,
husk, root, leaf). Unidirectionally cloned."
BASE COUNT      166 a      160 c      139 g      185 t
ORIGIN

```

```

alignment_scores:
  Quality:      75.50      Length:      94
  Ratio:        1.606      Gaps:        4
  Percent Similarity: 50.000  Percent Identity: 25.532

```

```

alignment_block:
US-10-048-197-2 x AW313326/rev ..

```

```

Align seg 1/1 to reverse of: AW313326 from: 1 to: 650

```

```

7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCysVa 23
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
425 AAGGCAACAGTCACAGTCCACAGAGGAGGCTTGGCATGGGT...GT 379
23 lSerThrGlyAsnValAlaMetIySglu...GlnAsnGlnGlnThrIleG 39
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
378 CGCTGTGCGCACTGGCGATGAGGAGAGAGCATGCCAGACATCC 329
39 lGlnThrIleIleIySglYsThrAsnIySglGlnIuIleSerSerArg 55
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
328 AGATG..... 324
56 PheGlySerAlaAspSerIleSerPheMetIleValIleIySPhel 72
  |::|::|::|::|::|::|::|::|::|::|::|::|::|
323 .....AGCGTCAACTTCCTTGTTCCTCTGTGAG..... 294
72 yHisThrAlaIleLeuAlaProAsnArgTyrGlnGlnIuIleLeuSerLeu 89
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
293 .....AAGAACTGCGAGAAATGTGAGATGCCCTGT 266
89 lIleSerPheLeuThrPValIySProTyrArg 99
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
265 ACGTCAAGAGCACCATGGGAGAGCATACAGG 234

```


1668 CCTTACAAAGGCGAGCTGATCGA 1693

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-13672-1

```
seq_documentation_block:
; Sequence 1, Application PC/TUS9513672
; GENERAL INFORMATION:
; APPLICANT: Meiser, Jeffrey M.
; TITLE OF INVENTION: Vaccines for Haemophilus influenza
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and
; ADDRESSEE: Norris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13672
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/332,576
; FILING DATE: October 31, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ralph, Rebecca L.
; REGISTRATION NUMBER: 35,152
; REFERENCE/DOCKET NUMBER: CH-536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1553..2005
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 209..1492
; PCT-US95-13672-1

alignment_scores:
Quality: 67.50 Length: 42
Ratio: 2.045 Gaps: 1
Percent Similarity: 78.571 Percent Identity: 33.333

alignment_block:
US-10-048-197-2 x PCT-US95-13672-1 ..
Align seg 1/1 to: PCT-US95-13672-1 from: 1 to: 2100

11 SerAlaLeuLeuThrAlaLeuValThGlyCysValSerThrGlyAs 27
1568 TCCTATTATTTTACCGCATTTTAATGATGCGCTGTGTGCAAAATGCGAA 1617
27 nValAlaMetLysGluGlnAsnGlnGlnThrIleGlu.....GlnI 41
1618 TGTAAACAACCTCAAGCGCAAAATGCAAGTAGAATAAGTGATTAAG 1667
41 hrIleIleLysGlyLysThrAsnLys 49
```

1668 CCTTACAAAGGCGAGCTGATCGA 1693

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-023-082A-23

```
seq_documentation_block:
; Sequence 23, Application US/09023082A
; Patent No. 6077692
; GENERAL INFORMATION:
; APPLICANT: RUBEN, STEVEN M.
; APPLICANT: JIMENEZ, PABLO
; APPLICANT: DUAN, D. ROXANNE
; APPLICANT: RAMPY, MARK A.
; APPLICANT: MENDRICK, DONNA
; APPLICANT: ZHANG, JDN
; APPLICANT: NI, JIAN
; APPLICANT: MOORE, PAUL A.
; APPLICANT: COLEMAN, TIMOTHY A.
; APPLICANT: GRUBER, JOACHIM R.
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: GENTZ, REINER L.
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,082A
; FILING DATE: 13-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01790
; FILING DATE: 14-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/461,195
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,852
; FILING DATE: 13-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,045
; FILING DATE: 28-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,432
; FILING DATE: 23-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/910,875
; FILING DATE: 13-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,561
; FILING DATE: 13-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0360008/EKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
```

MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 593..1216
US-09-023-082A-23

alignment_scores:
Quality: 67.50 Length: 100
Ratio: 1.298 Gaps: 4
Percent Similarity: 52.000 Percent Identity: 28.000

alignment_block:
US-10-048-197-2 x US-09-023-082A-23/rev ..

Align seg 1/1 to reverse of: US-09-023-082A-23 from: 1 to: 4177

```
19 ValThrglyCysValSerThrglyAsnValAlaMetlysgluInsnagl 35
|||||  ::|||  |||  |||:||||:|
699 GTGACAGGACGAGACACACAGAAACACAAACAGCAGCAGCA 650
|||  ::|  ::|  |||:||||:|
35 ngInThrIleGluInThrIleIlelysglyThrAsnlysgluInlur 52
|||  ::|  ::|  |||:||||:|
649 GCAGCGCGGCGAGGTGC.....GGAAGGCTGAGGCGACATGTG 612
|||  ::|  ::|  |||:||||:|
52 IeserSerArgPheglySerAlaAspSerIleSerPheMetIleValal 68
|||  ::|  ::|  |||:||||:|
611 TCAGTATTCATTTCACATGTGACGAACTCGGCGACTGGAATGTGC 562
|||  ::|  ::|  |||:||||:|
69 IlelyspPhegly.HisThrAlaIleLeuAlaPro..... 79
|||  ::|  ::|  |||:||||:|
* 561 TCATCGAAGACATCTGCGAAGGCTAGACCCGATGCAAGCAGCAG 512
|||  ::|  ::|  |||:||||:|
80 .....AsnArgTrpGluInlIleLeuSerLeuIleIleSerPheLeuTrp 94
|||  ::|  ::|  |||:||||:|
511 AGAGCTCGAGGTGTGG.....CTGCTGTTAGCTCCCTCTGG 474
|||  ::|  ::|  |||:||||:|
95 VallyserProtyrArgProlysaAsnLeuSerPheTyrlleuThrAlays 110
|||  ::|  ::|  |||:||||:|
473 GCCGGGATCTGCCAGAGTGAAGTGCACCAACATTCATCTCTGG 426
|||  ::|  ::|  |||:||||:|
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-988-856B-3

seq_documentation_block:

Sequence 3, Application US/0898856B
Patent No. 6291642
GENERAL INFORMATION:
APPLICANT: Weinstein, Jasaminder
TITLE OF INVENTION: NO. 6291642el Mammalian Cell Cycle Protein
FILE REFERENCE: 06843.0026-04: A-283 D
CURRENT APPLICATION NUMBER: US/08/988, 856B
NUMBER OF SEQ ID NOS: 31
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1700
TYPE: DNA
ORGANISM: Homo. sapiens
US-08-988-856B-3

alignment_scores:
Quality: 66.50 Length: 64
Ratio: 1.583 Gaps: 1
Percent Similarity: 65.625 Percent Identity: 26.562

alignment_block:
US-10-048-197-2 x US-08-988-856B-3/rev ..

Align seg 1/1 to reverse of: US-08-988-856B-3 from: 1 to: 1700

```
17 LeuLeuValThrIleGlyCysValSerThrglyAsnValAlaMetlysglu 33
|||||  |||  ::|  ::|  |||:||||:|
```

```
1181 TTGATCTGGGTGATGTCTGCAGAGAACCCAGCCACCCCTCTCCAGGAG 1132
|||||  |||  ::|  ::|  |||:||||:|
33 nAsngInGlnThrIleGluInThrIleIlelysglyThrAsnlysg 50
|||||  |||  ::|  ::|  |||:||||:|
1131 CACTAGCCACACATGTGACCAAGTATCAT.....ACCACACATG 1091
|||||  |||  ::|  ::|  |||:||||:|
```

```
50 IngIuIleSerSerArgPheglySerAlaAspSerIleSerPheMetIle 66
|||||  |||  ::|  ::|  |||:||||:|
1090 GCCAAATGTCTCATCTGGGCCCGCCAGCCGACACACTTCTGGCT 1041
|||||  |||  ::|  ::|  |||:||||:|
1040 GTGGCCATCAGTGTGGCCACATGTGTCTGTCTACCCGAC 999
|||||  |||  ::|  ::|  |||:||||:|
```

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-01806-3

seq_documentation_block:

Sequence 3, Application PC/TUS9501806
GENERAL INFORMATION:
APPLICANT: Weinstein, Jasaminder
TITLE OF INVENTION: Novel Mammalian Cell Cycle Protein
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc./Patent Operations/RBW
STREET: 1640 Dehavenland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/01806
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1700 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1616
PCT-US95-01806-3

alignment_scores:
Quality: 66.50 Length: 64
Ratio: 1.583 Gaps: 1
Percent Similarity: 65.625 Percent Identity: 26.562

alignment_block:
US-10-048-197-2 x PCT-US95-01806-3/rev ..

Align seg 1/1 to reverse of: PCT-US95-01806-3 from: 1 to: 1700

```
17 LeuLeuValThrIleGlyCysValSerThrglyAsnValAlaMetlysglu 33
|||||  |||  ::|  ::|  |||:||||:|
1181 TTGATCTGGGTGATGTCTGCAGAGAACCCAGCCACCCCTCTCCAGGAG 1132
|||||  |||  ::|  ::|  |||:||||:|
33 nAsngInGlnThrIleGluInThrIleIlelysglyThrAsnlysg 50
|||||  |||  ::|  ::|  |||:||||:|
1131 CACTAGCCACACATGTGACCAAGTATCAT.....ACCACACATG 1091
|||||  |||  ::|  ::|  |||:||||:|
50 IngIuIleSerSerArgPheglySerAlaAspSerIleSerPheMetIle 66
|||||  |||  ::|  ::|  |||:||||:|
1090 GCCAAATGTCTCATCTGGGCCCGCCAGCCGACACACTTCTGGCT 1041
|||||  |||  ::|  ::|  |||:||||:|
```

67 ValValIleLysPheGlyHisThrAlaIleLeuAlaProAsn 80
 |||::: |||::: |||::: |||:::
 1040 GTGGACACACTGTGGCCATGAGTTCCTACCGCAAC 999

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-196-520-5

seq_documentation_block:
 ; Sequence 5, Application US/09196520
 ; Patent No. 6204039
 ; GENERAL INFORMATION:
 ; APPLICANT: Falco, Carl S.
 ; APPLICANT: Allen, Stephen M.
 ; TITLE OF INVENTION: Plant Isocitrate Dehydrogenase Homologs
 ; FILE REFERENCE: BB-1291
 ; CURRENT APPLICATION NUMBER: US/09/196,520
 ; CURRENT FILING DATE: 1998-11-19
 ; EARLIER APPLICATION NUMBER: 60/067,388
 ; EARLIER FILING DATE: December 2, 1997
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5
 ; LENGTH: 1402
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; US-09-196-520-5

alignment_scores:
 Quality: 66.00 Length: 109
 Ratio: 1.100 Gaps: 6
 Percent Similarity: 55.046 Percent Identity: 26.606

alignment_block:
 US-10-048-197-2 x US-09-196-520-5 ..

Align seg 1/1 to: US-09-196-520-5 from: 1 to: 1402

* 15 ThrAlaLeuLeuValThrGlyCysValSerThrGlyAsnValAlaMet.. 30
 |||::: |||::: |||::: |||:::
 993 ACTGCTTACTCTGTAAGTGTGTACAAATGTCGCCCATTTACACTCA 1042
 31LysGluGlnAsnGlnGlnThrIleGluGlnThrIleLysG 45
 |||::: |||::: |||::: |||:::
 1043 TGACAAACAGCAGACATCAAAAGCCATCCTGAACAAATTCAGCAAG 1092
 45 jLysThrAsnLysGlnGlnLysSerArgPheGlySerLalaSper 61
 |||::: |||::: |||::: |||:::
 1093 GGAAGTACCGAAGCTGACCTCGGT.....GGCAGTTCAAGACA 1133
 62 lIleSerPheMetIleValValIleLysPheGlyHisThrAlaIleLeuAl 78
 |||::: |||::: |||::: |||:::
 1134 ACTGATTCACAAAGCAATATTT.....GATCATCTTTACATTTTG.. 1175
 78 aProAsnArgTrpGlnGluIleLeuSer..... 87
 |||::: |||::: |||::: |||:::
 1176GTTTGGAGTCACAGCAAAATTCCTCTTAT 1206
 88LeuIleIleSerPheLeuTrpValLysProTyraArgProLys 101
 |||::: |||::: |||::: |||:::
 1207 GATTTGATTTTCAGTCAATTTTCTATTATTAACCAATTAATCCGAAA 1256
 102 AsnLeuSerPheTyrlleuThrAlaLys 110
 |||::: |||::: |||::: |||:::
 1257 GAGTGGGAT..TACCTCAGCTTAAG 1280

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US96-05320A-333

seq_documentation_block:
 ; Sequence 333, Application PC/TUS9605320A
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences
 ; APPLICANT: 9410 Key West Avenue
 ; APPLICANT: Rockville, MD 20850

APPLICANT: United States of America
 APPLICANT: Johns Hopkins University
 APPLICANT: 720 Rutland Avenue
 APPLICANT: Baltimore, MD 21205
 APPLICANT: United States of America
 APPLICANT: Mark D. Adams
 APPLICANT: Owen White
 APPLICANT: Hamilton O. Smith
 APPLICANT: J. Craig Venter
 TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome
 NUMBER OF SEQUENCES: 48
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 STREET: 1100 New York Avenue, Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20003-3934
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/05320A
 FILING DATE: April 22, 1996
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/476,102
 FILING DATE: June 7, 1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/487,429
 FILING DATE: June 7, 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Eric K. Steffe
 REGISTRATION NUMBER: 36,688
 REFERENCE/DOCKET NUMBER: 1488.014PC01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 333:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 402 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 PCT-US96-05320A-333

alignment_scores:
 Quality: 65.50 Length: 42
 Ratio: 2.047 Gaps: 1
 Percent Similarity: 76.190 Percent Identity: 33.333

alignment_block:
 US-10-048-197-2 x PCT-US96-05320A-333 ..

Align seg 1/1 to: PCT-US96-05320A-333 from: 1 to: 402

11 SerAlaLeuLeuThrAlaLeuValThrGlyCysValSerThrGlyAs 27
 |||::: |||::: |||::: |||:::
 16 TCTCTTATTTTACCGCACTTTTATGACTGCTGTGCAAAAGCGAA 65
 27 nValAlaMetLysGlnGlnAsnGlnGlnThrIleGlu.....GlnT 41
 |||::: |||::: |||::: |||:::
 66 TGTAAACAACCTCAAGCGCAAAATGCAAGTAAAGTGAATTAAG 115
 41 hrIleIleLysGlyLysThrAsnLys 49
 |||::: |||::: |||::: |||:::
 116 CTTTCAAAAAGCGAGCTGTGTCGA 141

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-371-377-16

```

seq_documentation_block:
; Sequence 16, Application US/08371377
; Patent No. 5851764
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: DEVELOPMENT OF DNA PROBES AND
; TITLE OF INVENTION: IMMUNOLOGICAL REAGENTS SPECIFIC FOR CELL SURFACE-EXPRESSED
; TITLE OF INVENTION: MOLECULES AND TRANSFORMATION-ASSOCIATED GENES
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/371,377
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/37590-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 621..1817
; US-08-371-377-16

alignment_scores:
Quality: 64.50 Length: 119
Ratio: 1.057 Gaps: 5
Percent Similarity: 51.261 Percent Identity: 22.689

alignment_block:
US-10-048-197-2 x US-08-371-377-16 ..
Align seg 1/1 to: US-08-371-377-16 from: 1 to: 2128

18 leuValThrGlyCys.....ValSerThrGlyAsnValAl 29
|||:::|||||
646 TTGATATCTCTCTGTGAAATTTGAGACGACGACGACTACTATGATGACATTC 695
29 aMetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLysGlyL 46
:::|||||:::|||||
696 ATTGATGCCCGACGACAGACAGACTTATCCAAACATGATTACAGGAC 745
46 ys.....ThrAsn 48
:::|||||
746 CTCACAGGCTGACGTGCTGCTGATTTGTCGCTGCTGCTGTTGAT 795
49 LysGlnGluIleSerSerArgPheGlySerAlaAspSerIleSerPhe 65
:::|||||:::|||||
796 TTGAAGCTGTGATCTCCAAAGAAATGGGACAGACCGACGACATGCCCTTCTG 845

```

```

65 tile.....ValValIleLysPhe. 71
:::|||||
846 GCTTACACACTGGTGTGAACAACATAATTGCGTGTACAAATGGA 895
72 GlyHisThrAlaIleLeuAlaPro.AsnArgTrpGlnGluIleLeuSerL 88
||| ||| ||| ||| ::|||:::|||||:::|||||
896 TTCCATCGAGCCACCCTACAGCCAGAGAGATATGAGAAATTTGTAAG 945
88 euIleIleSerPheLeuTrpValLysProTyrArgProLysAsnLeuSer 104
::: :::::|||||:::|||||
946 AAGTCAGACACTTACATTAAAGAAATTTGGCTACAAACCCGACACAGTACA 995
105 Phe 105
|||
996 TTT 998

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-07-895-300A-2
seq_documentation_block:
; Sequence 2, Application US/07895300A
; Patent No. 5279823
; GENERAL INFORMATION:
; APPLICANT: Frenz, John
; APPLICANT: Shire, Steven J.
; APPLICANT: Sliwowski, Mary B.
; TITLE OF INVENTION: PURIFIED FORMS OF DNase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: palin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/895,300A
; FILING DATE: 19920608
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hensley, Max D.
; REGISTRATION NUMBER: 27,043
; REFERENCE/DOCKET NUMBER: 747
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1039 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-895-300A-2

alignment_scores:
Quality: 63.50 Length: 124
Ratio: 0.977 Gaps: 6
Percent Similarity: 52.419 Percent Identity: 24.194

alignment_block:
US-10-048-197-2 x US-07-895-300A-2 ..
Align seg 1/1 to: US-07-895-300A-2 from: 1 to: 1039

```

```

: NAME: Johnston, Sean A.
: REGISTRATION NUMBER: 35, 910
: REFERENCE/DOCKET NUMBER: P0747C4
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/425-3562
: TELEFAX: 415/952-9881
: TELERX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1039 base pairs
: TYPE: Nucleic Acid
: STRANDEDNESS: Single
: TOPOLOGY: Linear
: US-08-458-367-2

alignment_scores:
    Quality: 63.50      Length: 124
    Ratio: 0.977      Gaps: 6
    Percent Similarity: 52.419      Percent Identity: 24.194

alignment_block:
US-10-048-197-2 x US-08-458-367-2 ..

Align seg 1/1 to: US-08-458-367-2 from: 1 to: 1039

3 ASNHSHSHSLIEARGLETHRILLESERALEUDEUTHRALALEUVA 19
: :::::::::::::::::::: ||||| ||||||:::
145 CACCATCATCTCAGATATGAGGGGCGATGAGCTGCTGGGGGCGCTGGCC 194
19 lthrglycysvalserthrglyasnvalalametlys.....gluina 34
: : ::::: ||| |||:::||||
195 AGTGCGCGCCCTACGACGAGGCGCGCTGTCCTCGAAGATCCGACCTTCA 244
34 snglglthrlegluinthrllethrllethrglylthrasnlygln 50
: || |||||::: ::||| ::::: ||| :::
245 ACATCCACACATTGGGGGAGACCAAGATGTCCAATGCCACCCTGTGACG 294
51 .....gluileserserargphleglyseralaasperielserph 64
: ::||| |||||::: ||
295 TACATTGTGCAGATCTGAGCGCGCATGACATCCGC..... 330
64 emettilevallielysphegilyhstfhalalelu..... 77
: ::::: ::|||:::
331 CTGTGCAGGAGGTCAGAGACGACGACCTGACTGCCGTGGGCAAGCTGC 379
78 .....alaproasnargtrpgingluileuser 87
380 TGGACACCTCATATGAGGATGCACACAGACCTTCACTACGTAGTGTGACG 429
88 leuileileser.....phleutrpvallyspr 97
: ::::: ::|||:::
430 GAGCCACGTGGGAGCGAGACGCTATTAAGAGCGCTACCTGTTGCG.... 474
97 GTYATRPProlysnleuser 104
: ||||||| ::::: ||
475 TACAGCGCTTACCAAGGTGTCT 495

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-05136-2-2-2
seq_documentation_block:
: Sequence 2, Application PC/TU9305136
: GENERAL INFORMATION:
: APPLICANT: Genentech, Inc.
: TITLE OF INVENTION: PURIFIED FORMS OF DNase
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080

```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 Inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: palin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05136
; FILING DATE: 19930528
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Johnston, Sean A.
; REGISTRATION NUMBER: 935,910
; REFERENCE/DOCKET NUMBER: 747PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3562
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1039 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
;
PCT-US93-05136-2

alignment_scores:
  Quality: 63.50      Length: 124
  Ratio: 0.977       Gaps: 6
  Percent Similarity: 52.419   Percent Identity: 24.194

alignment_block:
US-10-048-197-2 x PCT-US93-05136-2  ..

Align seg 1/1 to: PCT-US93-05136-2 from: 1 to: 1039

3 AsnHSHSHSLIeArlgLeuThrlleSerAlaLeuThrlAlaLeuYuA 19
  :::::::::::::::::::: ||||| ||||| ::::::::::::::
145 CACCATCATCTCAGATCGAGGCGCATGAAGCTGCGGGCGCTGCGC 194
19 lthrglCysValserThrglYasnValAlaMetlys.....GlulGlnA 34
  : :::::::::::::: ||| ||| :::::::::::::: |
195 ACTGGGGGCCCTACTGCAAGGGGCCGTGCTCGTAAGATGCGAGCCTTCA 244
34 snglGlnThrlleGluGlnThrllellelysglyLysThrsAnlysgln 50
  || ||||| :::::::::: :::::::::: ::|
245 ACATCCAGACATTGGGCGAGACCAAGATGCCAATGCCACCTCGTCAGC 294
51 .....GlulIeserAtrpPheglySerAlaAspSerIleSerph 64
  : ::||| ||||| ::|
295 TACATTGTGCAGATCTCTGAGCCGCTATGACATCGCC..... 330
64 eueItleValIlelysrPheglyHlstrAlaIleleu..... 77
331 CTGTGCCAGAGTCAGACAGACACCTGACTGCGCGTGGGGAAGCTGC 379
78 .....AlaProAsnAtrpGlnGlnIleLeuSer 87
  : :::::::::: ::|
380 TGGACAACTCAATCAGATGCGACCAACCACTATCAGTACGTGTCAGT 429
88 LeuIleIleSer.....PheLeuTrpValIlySPr 97
  : ::||| ::|
430 GAGCCACTGGAGGGAACAGTATAAGAGCGCTACCTGTCTGCTG..... 474
97 oTyrrAtrpProlysaAnLeuSer 104
  ||||| ::::::::::
475 TACAGGCTTGACACAGGTGCT 495

seq_name: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq: PCT-US93-10519-1
```

```

seq_documentation_block:
; Sequence 1, Application PC/TUS9310519
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; TITLE OF INVENTION: Compaction Assay for Assessment of Respiratory Disease
; TITLE OF INVENTION: Therapy
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 Inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: palin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10519
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Johnston, Sean A.
; REGISTRATION NUMBER: 35,910
; REFERENCE/DOCKET NUMBER: 792C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3562
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1039 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
PCT-US93-10519-1

alignment_scores:
  Quality: 63.50      Length: 124
  Ratio: 0.977       Gaps: 6
  Percent Similarity: 52.419   Percent Identity: 24.194

alignment_block:
US-10-048-197-2 x PCT-US93-10519-1  ..

Align seg 1/1 to: PCT-US93-10519-1 from: 1 to: 1039

3 AsnHSHSHSLIeArlgLeuThrlleSerAlaLeuThrlAlaLeuYuA 19
  :::::::::::::::::::: ||||| ||||| ::::::::::::::
145 CACCATCATCTCAGATCGAGGCGCATGAAGCTGCTGCGGGCGCTGCGC 194
19 lthrglCysValserThrglYasnValAlaMetlys.....GlulGlnA 34
  : :::::::::::::: ||| ||| :::::::::::::: |
195 ACTGGGGGCCCTACTGCAAGGGGCCGTGCTCGTAAGATGCGAGCCTTCA 244
34 snglGlnThrlleGluGlnThrllellelysglyLysThrsAnlysgln 50
  || ||||| :::::::::: :::::::::: ::|
245 ACATCCAGACATTGGGCGAGACCAAGATGCCAATGCCACCTCGTCAGC 294
51 .....GlulIeserAtrpPheglySerAlaAspSerIleSerph 64
  : ::||| ||||| ::|
295 TACATTGTGCAGATCTGAGCCGCTATGACATCGCC..... 330
64 eueItleValIlelysrPheglyHlstrAlaIleleu..... 77
331 CTGTGCCAGAGTCAGACAGACACCTGACTGCGCGTGGGGAAGCTGC 379
```

```

78 .....AlaProAsnArgTrpGlnGluIleLeuSer 87
      |||||: : : : :
380 TGGACAACCTCATCGATGCACGACAGACCTATCAGCTGCGCACT 429
      88 LeuIleIleSer.....PheLeuTrpValIysPfr 97
      : : : : :
430 GAGCCACTGGAGCGAAGAGCTATAGAGCGCTACCTGCTGTG..... 474
      97 oTyrrArgProIysAsnLeuSer 104
      |||||: : : : :
475 .TACAGCGCTGACAGAGTGTCT 495

```

seq_name: /cgn2_6/plodata/2/ina/6B_COMB.seq:US-09-013-810-1

```

seq_documentation_block:
; Sequence 1, Application US/09013810
; Patent No. 6197551
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: No. 6197551el TANGO 80 Molecules and Uses Thereof
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/013,810
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MEI-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 746 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 135..428
; US-09-013-810-1

```

```

alignment_scores:
  Quality: 63.00      Length: 111
  Ratio: 1.050      Gaps: 5
  Percent Similarity: 54.054      Percent Identity: 24.324

```

alignment_block:

```

US-10-048-197-2 x US-09-013-810-1 ..
Align seg 1/1 to: US-09-013-810-1 from: 1 to: 746

```

```

4 HistHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuValIth 20
  |||||: : : : :
117 CATCACTTGGCTTACAGATGTTAGAGATCTTA.....GTAGTCGTGTG 160

```

seq_name: /cgn2_6/plodata/2/ina/6A_COMB.seq:US-09-058-489-34

```

seq_documentation_block:
; Sequence 34, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WH197-08PA
; CURRENT APPLICATION NUMBER: US/09/058,489
; EARLIER FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 10091
; TYPE: DNA
; ORGANISM: Human
; US-09-058-489-34

```

```

alignment_scores:
  Quality: 63.00      Length: 103
  Ratio: 1.086      Gaps: 4
  Percent Similarity: 56.311      Percent Identity: 23.301

```

alignment_block:

```

US-10-048-197-2 x US-09-058-489-34 ..
Align seg 1/1 to: US-09-058-489-34 from: 1 to: 10091

```

```

10 IleSerAlaLeuLeuThrAlaLeuValIthGlyCysValSerThrGI 26
  ||| |||: : : : :
5382 ATTAGGCTTATACAGAAATATATCTGGCATCGCATGCGGGGCAATTAGG 5431
      26 yAsnValAlaMetIysGluGlnAsnGlnInThrIleGlnInThrIleI 43
      | ||| : : : : :
5432 ACTAGTTTATGAGCAATGAGAAATTAACCTAAATTTTACAGATGACCA 5481
      43 IeysGlyIysThrAsnIysGlnIleIleSerSerArgPheIysSerAla 59
      : : : : :
5482 CCAATGA...AGCAATTAAGCTGAGAGGTGGAAGATCAACAAGTTGCTGT 5528
      60 AspSerIleSerPheMetIleValValIleIysPheGlyHisThrAlaI 76

```



```

5529 GAAGCTGAGTGTGACCTTATGTTT.....GC 5560
76 eLeuAlaPro.AsnArgTrpGlnGluIleuSerIleuSerPhe 92
5561 TTACTTCCAGACGGCTGTGATGACTAGTAAG..... 5595
93 LeuTrpValIysProTyrArgProIysAsnLeuSerPheTyrLeuThrAl 109
5596 .....AAAAGCCTGGACAGACCTTCATCATCTTATTATTCACACTGT 5639
109 aLysAla 111
5640 CCAAGCA 5646

```

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-366-783-3

seq_documentation_block:

```

: Sequence 3, Application US/08366783
: Patent No. 5650554
: GENERAL INFORMATION:
: APPLICANT: Moloney, Maurice M
: TITLE OF INVENTION: Oil-Body Proteins As Carriers of
: NUMBER OF INVENTION: High-Value Peptides in Plants
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DEHLINGER & ASSOCIATES
: STREET: 350 CAMBRIDGE AVENUE, SUITE 250
: CITY: PALO ALTO
: STATE: California
: COUNTRY: United States
: ZIP: 94025-1536
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/366,783
: FILING DATE:
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: FABIAN, GARY
: REGISTRATION NUMBER: 33,875
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-324-0880
: TELEFAX: 415-324-0960
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 765 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-366-783-3

```

alignment_scores:

Quality:	61.50	Length:	98
Ratio:	1.098	Gaps:	4
Percent Similarity:	57.143	Percent Identity:	23.469

alignment_block:

US-10-048-197-2 x US-08-366-783-3 ..

Align seg 1/1 to: US-08-366-783-3 from: 1 to: 765

```

8 LeuThrIleSerAlaLeuLeuThr...AlaLeuLeuValThrGlyCysVa 23
: : : : : : : : : : : : : : : : : : : : : : : : : : :
460 ATCTCTGTCGGCTCTCATACCGTAGCACTTCTCATCTGCTTCTTCT 509
23 IserThrGlyAsnValAlaMet..... 30
: : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

510 CTCCTCTGTGGTGGTTTGCATTGACGCTATTAACCGTCTCTCCGTGATCT 559
31 .....LysGluGlnAsnGlnIleuThrIleGluGln 40
560 ATTAGTACCCAAACGGAGACACCCACAGGGCTCAATATGTCACACT 609
41 ThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg..PheG 57
: : : : : : : : : : : : : : : : : : : : : : : : : : :
610 GCAGAGATGATAGCTGGACCAACAGCTCAGATATTAAGACAGACTCA 659
57 IysSerAlaAspSer.....IleSerPheMetIleValIle 69
: : : : : : : : : : : : : : : : : : : : : : : : : : :
660 ATACTGACGACAGACATACATGAGTGTGTCAGCATGACCGTACCTACTC 709
70 LysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 83
710 GTGGTGGCCAGCACACTACTATTCGAAGAGAACCTGGCGC 751

```

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-514-302-1

seq_documentation_block:

```

: Sequence 1, Application US/09514302
: Patent No. 6338959
: GENERAL INFORMATION:
: APPLICANT: HATADA, Yuji
: APPLICANT: IGARASHI, Kazuaki
: APPLICANT: OZAKI, Katsuya
: APPLICANT: ARA, Katsutoshi
: APPLICANT: KAWAI, Shuji
: APPLICANT: ITO, Susumu
: TITLE OF INVENTION: GENE FOR ENZYME HAVING BOTH ALKALINE PHOSPHATASE AND
: FILE REFERENCE: 2173-105P
: CURRENT APPLICATION NUMBER: US/09/514,302
: CURRENT FILING DATE: 2000-02-28
: EARLIER APPLICATION NUMBER: 08/952,084
: EARLIER FILING DATE: 1997-11-10
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 1
: LENGTH: 6142
: TYPE: DNA
: ORGANISM: Bacillus sp.
: NAME/KEY: CDS
: LOCATION: (145)..(5958)
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: (145)..(240)
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: (241)..(5958)
: US-09-514-302-1

```

alignment_scores:

Quality:	61.50	Length:	113
Ratio: <td>1.118 <td>Gaps: <td>4 </td></td></td>	1.118 <td>Gaps: <td>4 </td></td>	Gaps: <td>4 </td>	4
Percent Similarity: <td>48.673 <td>Percent Identity: <td>21.239 </td></td></td>	48.673 <td>Percent Identity: <td>21.239 </td></td>	Percent Identity: <td>21.239 </td>	21.239

alignment_block:

US-10-048-197-2 x US-09-514-302-1 ..

Align seg 1/1 to: US-09-514-302-1 from: 1 to: 6142

```

4 HisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValTh 20
: : : : : : : : : : : : : : : : : : : : : : : : : : :
5672 CACCACACAGGAGATGGGATG...GGGATGAAACACGGCTCTC..... 5713
20 rGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnIle 37
: : : : : : : : : : : : : : : : : : : : : : : : : : :
5714 .....CAGGCAACGTAATGCAATTAATCCAGAACACACAGGAAA 5756

```

```

alignment_scores:
  Quality: 61.00      Length: 124
  Ratio: 0.910      Gaps: 5
  Percent Similarity: 54.032      Percent Identity: 23.387

alignment_block:
US-10-048-197-2 x US-09-309-487-25/rev ..

Align seg 1/1 to reverse of: US-09-309-487-25 from: 1 to: 25533

2 LeuAnshHsHsIleArgLeuThrIleSerAlaLeuDeuThrAlaLeuIle 18
  GTCAACACGCAAGAAAGCCAGAGACCCCAAGCCCTTTATACACTGTGGCAA 2412
  ::::::::::: ||| ||| :::::::::::||||:::
18 uValThrGlyCysValSerThrGlyAsnValAlaMet.....L 31
  :::: ||| :::::::::::||||:::
2362 ACTCTCTGCTGTGCMAAGCAGCACTCAACCTCTGCTGACTCTGGGACAC 2113
  31 ysgIugInaIngInGlnThrIleGluInGlnThrIleIleGlyIyLyrThr 47
  AGACAAAGACATCATAAATGATGACGAAGGCTCGACGTGAGATGACAAAGGG 2212
  ::::::::::: ||| ||| :::::::::::
48 AsnYsgInGlnIleSerSer.....Ar 55
  ||| |||:::
2262 APhCTCTGGAAAGATCATGTCTCTTAATGAGGAGATGCTAGCTGCATTTGG 2213
  55 pHeCTYSerAlaAspSerIleSer.....PheMetIleV 67
  | ||||| :::::::::::|||| ::::
2212 GGGCGGTAGCATATACACACTCTCACTAGTAGGAATTAATACACAGACTG 2163

```

```

67 yAValIlelYSPHGIyHisPhrAlaleLeuAlProAsnArgTrpIn 83
||| |||||||:::
2162 TGCt.....GCTCACACAGATT ..... 2144

      84 GLuileUeSrLleUlleiSerPheLeUTrpVAllySPrOTryArGr 100
         || |   ||| :::::::::::::::::::: ||| :::
    2143 GAgCttATTCtTACTTTGTCCctCATTTTTTGcCGccCAACA 2094
          ||:::||::: |||
        100 OLySnleUsErPHeTyrlEU 107
           ||:::||::: |||
       2093 CAAGAACCtgAAgcAcTCTTA 2072

seq_name:/cgcn2_6/pldata/2/lna/6A.COMB.seq:US-08-878-474-6

seq_documentation_block:
? Sequence 6, Application US/08878474
? Patent No. 6133232
? GENERAL INFORMATION:
? APPLICANT: De Robertis, Edward M.
? Applicant: Boummeester, Tewls
? TITLE OF INVENTION: Endoderm, Cardiac and Neural Inducing
? TITLE OF SEQUENCE: Factors
? NUMBER OF SEQUENCES: 10
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Majestic, Parsons, Siebert & Hsu
? STREET: Four Embarcadero Center, Suite 1100
? CITY: San Francisco
? STATE: California
? COUNTRY: U.S.A.
? ZIP: 94111-4106
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patelnin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878.474
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/020,150
FILING DATE: 20-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Siebert, J. Suzanne
REGISTRATION NUMBER: 28,758
REFERENCE/DOCKET NUMBER: 3100,.002U51
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/248-5500
TELEX: 415/362-5418
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3655 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-878-474-6

alignment_scores:
Quality: 61.00 Length: 100
Ratio: 1.034 Gaps: 3
Percent Similarity: 59.000 Percent Identity: 23.000

alignment_block:
US-10-048-197-2 x US-08-878-474-E/rev ..

Align seg 1/1 to reverse of: US-08-878-474-6 from: 1 to: 36555

3 AsntHSISILeAtgLeUrThrLESerAlAleuLeUrHrAlAleuLeuVa 19
|||||:::|::: ||| ||||| ||| :|||:::
1367 AACATCATGGTAGAACGCCATAAGCTTGCGTAGCTTAAAGTCGCAT 1318
```



```
Align seg 1/1 to reverse of: US-09-449-083-14 from: 1 to: 507
7 ArgLeuThrIleSerAlaLeuThrAlaLeuValThrGlyCysVa 23
||||| ||||| ||| |||:|||||:|||||:|||||
192 AGACTGAAGATTTCGAAACTCTGCTGCTGATGTTGACCTGCGCT 143
23 lSerThrGlyAsnValAlaMetLysGlnGlnAsnGlnGlnThrIleGlu 40
|||||:|||||: ||| |||:||||| ||||| :|||
142 CGCGACCGGCTCT...CCCTACGCGGAAACACGCGCAGACTACCAATG 96
40 lnrThr 41
95 AAAGC 91

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-868-373-3
seq_documentation_block:
; Sequence 3, Application US/08868373
; Patent No. 6307128
; GENERAL INFORMATION:
; APPLICANT: Jaworski, Jan G.
; APPLICANT: Post-Belttenmiller, Martha A.
; TITLE OF INVENTION: FATTY ACID ELONGASES
; FILE REFERENCE: 07148/064001
; CURRENT APPLICATION NUMBER: US/08/868,373
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-08-868-373-3

alignment_scores:
Quality: 59.50 Length: 87
Ratio: 1.214 Gaps: 4
Percent Similarity: 56.322 Percent Identity: 22.989

alignment_block:
US-10-048-197-2 x US-08-868-373-3 ..

Align seg 1/1 to: US-08-868-373-3 from: 1 to: 1479
22 CysValSerThrGlyAsnValAlaMetLysGlnGlnAsnGlnGlnThrI 38
|||||:|||||: |||||: |||||: |||||: |||||:
1156 TGTATCCATCGCGGTGCTAGAGCGCTAATTGATGAGATGGAAGAATCT 1205
38 eGlnGlnThrIleIleLysGlyLysThrAsnLysGlnGlnIleSerSer 55
||||| ||| :||| :||| :||| :||| :||| :|||
1206 TCATCTACTCCACTAGACGCTTGAGGCTTCAGAGATGACATTA...CACA 1252
55 rrpPheGlySerAlaAspSerIleSerPhe..... 64
|||||:|||||: |||||: |||||: |||||: |||||:
1253 GGTGTGTATACCTCTTCGAGCTCCATTTGTAAGATGATGCGATTCACA 1302
65 .....MetIle.ValValIleLysPheGlyHisThrAlaI 76
||||| ||| ||| |||||: |||
1303 GAAGCCAAAGAGATGACGAAAGAGATGATGCGATGCGATGCGATGCG 1348
76 lLeuAlaProAsnArgTrpGlnGlnIleLeuSerIleIleSerPhe 92
|||||:|||||: |||||: |||||: |||||: |||||:
1349 .....CGTTGGGCTCAGGTTTAAAGTAAATGATTCAGTTT 1384
93 LeuTrpVal 95
|||||:
1385 GGTGCGCTC 1393

seq_name: /cgn2_6/ptodata/2/lna/Backfile1.seq:5225348-2
seq_documentation_block:
```

```
; Patent No. 5225348
; APPLICANT: HAGATA, SHIGEKAZU; SUGANO, SUMIO; KIM, DONG W.;
; IETSUKI, FAICHI; KAZIRO, YOSHITO
; TITLE OF INVENTION: DNA FRAGMENT AND EXPRESSION PLASMID
; CONTAINING THE DNA FRAGMENT
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/447,823
; FILING DATE: 08-DEC-1989
; SEQ ID NO: 2
; LENGTH: 1753
5225348-2

alignment_scores:
Quality: 59.50 Length: 127
Ratio: 0.902 Gaps: 5
Percent Similarity: 51.969 Percent Identity: 21.260

alignment_block:
US-10-048-197-2 x 5225348-2 ..

Align seg 1/1 to: 5225348-2 from: 1 to: 1753
10 lIleSerAlaLeuThrAlaLeuValThrGlyCys..... 22
|||||: |||||: |||||: |||||: |||||: |||||:
256 CTGACGCTGAACGTGATCACCATTGATCTCTGTAATTTGAG 305
23 ...ValSerThrGlyAsnValAlaMetLysGlnGlnAsnGlnGlnThrI 38
||||| ||||| :|||||: |||||: |||||: |||||:
306 ACCAGAGATCATGTGACTATCATTTGATGATGCCAGACAGACAGACTT 355
38 lGlnGlnThrIleIleLysGlyLys..... 46
|||||: |||||: |||||: |||||: |||||: |||||:
356 TATCAAAATCATGATTCACAGGACATCTCAGCTGCTGCTCTCTGA 405
47 .....ThrAsnLysGlnGlnIleSerSerArgPheG 57
|||||: |||||: |||||: |||||: |||||: |||||:
406 TTGTTGCTGCTGCTGTTGTAATTTGAAGCTGTTATCTCAGAAATGG 455
57 ySerAlaAspSerIleSerPheMetIle..... 66
|||||: |||||: |||||: |||||: |||||: |||||:
456 CAGACCCGAGAGCATGCCCTTCGCTTACACACTGGGTGTAACAACACT 505
67 .....ValValIleLysPhe.GlyHisThrAlaIleLeuAlaProA 80
|||||: |||||: |||||: |||||: |||||: |||||:
506 AATTGTCGCTTACAAATGATTCACCTGACGCCACCTACAGCCAGA 555
80 snArgTrpGlnGlnIleLeuSerLeuIleIleSerPheLeuTrpValLys 96
|||||: |||||: |||||: |||||: |||||: |||||:
556 AGAGATATGAGAAATTTGTAAGAGAGTCACTTACATTAGAAATTT 605
97 ProTyrArgProLysAsnLeuSerPhe 105
|||||: |||||: |||||: |||||: |||||: |||||:
606 GGCTACAAACCCGACACAGTACGATTTT 632

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-484-101B-49
seq_documentation_block:
; Sequence 49, Application US/08484101B
; Patent No. 5824868
; GENERAL INFORMATION:
; APPLICANT: California Institute of Technology
; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard F. Trecairtin
; STREET: 3400 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,101B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/
; FILING DATE: 01-JUL-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/086,555
; FILING DATE: 01-JUL-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Tregearlin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-57515-2/RPT
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2405 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 288..2196
; US-08-484-101B-49
```

```
alignment_scores:
  Quality: 59.50      Length: 84
  Ratio: 1.526        Gaps: 2
  Percent Similarity: 46.429  Percent Identity: 23.810
```

alignment_block:

US-10-048-197-2 x US-08-484-101B-49 ..

Align seg 1/1 to: US-08-484-101B-49 from: 1 to: 2405

```
22 CysvalSerThrGlyAsnValAlaMetLysGluGlnAsnGlnInthr.. 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
195 TGTGTGCGACGACGAAATAGCATCTTAAAGGAAGAAAGAACTCG 244
37 .....
245 AAGTTACTAAAAATTTTGATCTTTGGGACGAAACGAGATTAATGAAT 294
38 .....
295 CCGTGATTCATTCAGAGCTTTACTGCGCAACCTGTGACCTCGTGTTAA 344
49 LysGlnGluLeuSerArgPheGlySerAlaAspSerIleSerPhe 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 TACCAATACCTCTCAGATTCTTCATTTGCTGTAAGCTTTCATTTCA 394
65 ttlevalvallllelypheglyhsthrAlaIleleuAlaProAsnArgT 82
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 GTTGAAGCTTATTATTGTCGCAAAATCTGCATGCTTCCCATACGAT 444
82 rp 82
445 GG 446
```

seq_name: /cgn2.6/ptodata/2/1na/6B_COMB.seq:US-08-714-524D-49

seq_documentation_block:

```
; Sequence 49, Application US/08714524D
; Patent No. 6294716
; GENERAL INFORMATION:
; APPLICANT: Meyerowitz, Elliott M
; APPLICANT: Chang, Karen
; APPLICANT: Blecker, Anthony B
; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
; FILE REFERENCE: a-57515-4
; CURRENT APPLICATION NUMBER: US/08/714,524D
; CURRENT FILING DATE: 1996-09-16
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 49
; LENGTH: 2405
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (288)..(2195)
; US-08-714-524D-49
```

```
alignment_scores:
  Quality: 59.50      Length: 84
  Ratio: 1.526        Gaps: 2
  Percent Similarity: 46.429  Percent Identity: 23.810
```

alignment_block:

US-10-048-197-2 x US-08-714-524D-49 ..

Align seg 1/1 to: US-08-714-524D-49 from: 1 to: 2405

```
22 CysvalSerThrGlyAsnValAlaMetLysGluGlnAsnGlnInthr.. 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
195 TGTGTGCGACGACGAAATAGCATCTTAAAGGAAGAAAGAACTCG 244
37 .....
245 AAGTTACTAAAAATTTTGATCTTTGGGACGAAACGAGATTAATGAAT 294
38 .....
295 CCGTGATTCATTCAGAGCTTTACTGCGCAACCTGTGACCTCGTGTTAA 344
49 LysGlnGluLeuSerArgPheGlySerAlaAspSerIleSerPhe 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 TACCAATACCTCTCAGATTCTTCATTTGCTGTAAGCTTTCATTTCA 394
65 ttlevalvallllelypheglyhsthrAlaIleleuAlaProAsnArgT 82
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 GTTGAAGCTTATTATTGTCGCAAAATCTGCATGCTTCCCATACGAT 444
82 rp 82
445 GG 446
```

seq_name: /cgn2.6/ptodata/2/1na/6A_COMB.seq:US-09-058-489-19

seq_documentation_block:

```
; Sequence 19, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WH197-08PA
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
```

[illegible]

```

2714 ATACAGACCTGAACAGACAGACAGAAAGCCATGAGAAAAACAGATCTCT 2763
57 GYSerAlaAspSerIleSerPheMetIleValIleLeuSerLeuIle 73
:::|||||:.....:
2764 ACCTCGCTCAGAGGCGCTTCAT.....AAAAGTCAGAG 2795
73 sTrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIle 90
:::|||||:.....:
2796 TTCATCTTTTCAGACGACCTAATGACAGACAACTCTTTCCAC..... 2840
90 leSerPheLeuTrpValLysProTyrArgProLysAsnLeuSerPheTyr 106
|||||:.....:
2841 .....TGGTCAGCGCCAGTATCACAGCAACTAGCACTGGTAT 2879

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-07-961-522-3

seq_documentation_block:
; Sequence 3, Application US/07961522
; Patent No. 5417971
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: GERLACH, GERALD F.
; APPLICANT: WILSON, PHILIP J.
; APPLICANT: ROSSI-CAMPOS, AMALIA
; TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
; TITLE OF INVENTION: PLEUROPEINOMINAE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBERTA L. ROBINS
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/961,522
; FILING DATE: 19921015
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9000-0015.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1903 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1777
; US-07-961-522-3

alignment_scores:
Quality: 59.00 Length: 76
Ratio: 1.439 Gaps: 4
Percent Similarity: 53.947 Percent Identity: 27.632

alignment_block:
US-10-048-197-2 x US-07-961-522-3 ..
Align seg 1/1 to: US-07-961-522-3 from: 1 to: 1903

```

```

38 IleGluGlnThrIleIle.....LysGlyLysThrAsnLysGlnGluI 52
:::|||||:.....:
878 TTCAGGCAACCGTTTAGGGGAAGAACCAAGCAAGAAAAAGCATGAT 927
52 eSerArgPheGlySerAlaAspSerIleSerPheMetIleValIle 69
|.....:
928 CCGA.....TCTTACTGATTCAAATTA 950
69 leLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln..... 83
|||:.....:
951 TCCTGAAGGGGATTCATGCTCTTAAGCTGAAGAAATGCGAGGGAAT 1000
84 .....GluIleLeuSerLeuIleIleSerPheLeuTrpValLys..... 96
:::|||||:.....:
1001 TTTTCACAAATTAATAATCTCTCTTTCAGATATTTCGACGTAATAAGTGA 1050
97 ....ProTyrArgProLysAsnLeuSer 104
|||||:.....:
1051 AACGGCGAGAGACAGACAGAAAGCAATCA 1078

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-217-438-3

seq_documentation_block:
; Sequence 3, Application US/08217438
; Patent No. 5521072
; GENERAL INFORMATION:
; APPLICANT: Potter, Andrew A.
; APPLICANT: Gerlach, Gerald F.
; APPLICANT: Wilson, Philip J.
; APPLICANT: Rossi-Campos, Amalia
; TITLE OF INVENTION: ACTINOBACILLUS PLEUROPEINOMINAE
; TITLE OF INVENTION: TRANSFERRIN BINDING PROTEINS AND USES THEREOF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/217,438
; FILING DATE: 22-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9001-0015.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-617-8999
; TELEFAX: 415-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1903 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1777
; US-08-217-438-3

alignment_scores:
Quality: 59.00 Length: 76
Ratio: 1.439 Gaps: 4
Percent Similarity: 53.947 Percent Identity: 27.632

```

```

alignment_block:
US-10-048-197-2 x US-08-217-438-3 ..
Align seg 1/1 to: US-08-217-438-3 from: 1 to: 1903
38 IIEGLGLThrllele.....LysGlyLysThrasnLysGlnIuIl 52
      ::::: |||::: || ||| ::||| ||
878 TTCACGCAACCGCTTTAGGGGAAAGCCAAAGCAGAAAGAGAGGTGAT 927
52 eSerSerArgpHeGlySerAlaSpSerIleSerPheMetIleValI 69
928 CCGA.....TCTTACTGATTCAAATTA 950
69 IeLysPheGlyHsThrAlaIleLeuAlaProAsnArgTrpGln..... 83
951 TCTTGAGGGGGATTCATGCTCTAAAGCTGAAGAAATGGCAGGGAAGT 1000
84 .....GluIleuSerIleuIleIleSerPheLeuTrpValys..... 96
1001 TTTTCACAAATATAATCTCTTTCAGATATTTCAGCTAAAGTGAA 1050
97 ...ProTyrArgProLysAsnLeuSer 104
1051 AACGGCGAGACGACACAGACGAATCA 1078

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-321-978-3

seq_documentation_block:
; Sequence 3, Application US/08321978
; Patent No. 5801018
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: GERLACH, GERALD F.
; APPLICANT: WILSON, PHILIP J.
; APPLICANT: ROSSI-CAMPOS, AMALIA
; TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBERTA L. ROBINS
; STREET: 285 HAMILTON AVE, SUITE 200
; CITY: PALO ALTO.
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/321.978
; FILING DATE: 12-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9001-0015.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 327-3400
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1903 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1777
; US-08-321-978-3

```

```

alignment_scores:
Quality: 59.00 Length: 76
Ratio: 1.439 Gaps: 4
Percent Similarity: 53.947 Percent Identity: 27.632

alignment_block:
US-10-048-197-2 x US-08-321-978-3 ..
Align seg 1/1 to: US-08-321-978-3 from: 1 to: 1903
38 IIEGLGLThrllele.....LysGlyLysThrasnLysGlnIuIl 52
      ::::: |||::: || ||| ::||| ||
878 TTCACGCAACCGCTTTAGGGGAAAGCCAAAGCAGAAAGAGAGGTGAT 927
52 eSerSerArgpHeGlySerAlaSpSerIleSerPheMetIleValI 69
928 CCGA.....TCTTACTGATTCAAATTA 950
69 IeLysPheGlyHsThrAlaIleLeuAlaProAsnArgTrpGln..... 83
951 TCTTGAGGGGGATTCATGCTCTAAAGCTGAAGAAATGGCAGGGAAGT 1000
84 .....GluIleuSerIleuIleIleSerPheLeuTrpValys..... 96
1001 TTTTCACAAATATAATCTCTTTCAGATATTTCAGCTAAAGTGAA 1050
97 ...ProTyrArgProLysAsnLeuSer 104
1051 AACGGCGAGACGACACAGACGAATCA 1078

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-710-584-3

seq_documentation_block:
; Sequence 3, Application US/08710584
; Patent No. 5876725
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: GERLACH, GERALD F.
; APPLICANT: WILSON, PHILIP J.
; APPLICANT: ROSSI-CAMPOS, AMALIA
; TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBERTA L. ROBINS
; STREET: 285 HAMILTON AVE, SUITE 200
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710.584
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/321.978
; FILING DATE: 12-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9001-0015.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 327-3400
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

```


alignment_scores:			
Quality:	59.00	length:	73
Ratio:	1.666	Gaps:	1
Percent Similarity:	47.945	Percent Identity:	26.027
alignment_block:			
US-10-048-157-2	x	US-09-564-805-223	..
Align seg 1/1	to:	US-09-564-805-223	from: 1 to: 290

```

align -block:
US-10-048-197-2 x US-09-564-805-223  ..

Align seg 1/1 to: US-09-564-805-223 from: 1 to: 290

22 CysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnT
   ||| ||||| ||||| ||| |||||
1379 TGCAGAGAGTACAGAGAGAGTGCACAGAGAGGCCACCCACAGACAG
   ||| ||||| ||||| ||| |||||
38 eGluGlnThrIleLeuLysGlyLysThrAsnLysGlnGluIleSer

```

1429 AGAAGTCAGTACCCAGAATCATCTTCCITTGACACAGGGTCCATCCC 1478

—
—

—
—
—
..
..
..

..
..
..

4

[illegible][illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

; Sequence 59, Application US/08121713D

; GENERAL INFORMATION:

APPLICANT: KOLODKIN, Alex L.

APPLICANT: Bentley, David R.

1 TITLE OF INVENTION: The Semaphorin Gene Family

CORRESPONDENCE ADDRESS:

STREET: 268 Bush Street, Suite 3200

STATE: CA

ZIP: 94104

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

```

; CURRENT APPLICATION DATA:

```

FILING DATE: 13-SEP-1993

ATTORNEY/AGENT INFORMATION:

REGISTRATION NUMBER: 36,627

TELECOMMUNICATION INFORMATION:

TELEFAX: (415) 343-4342

```

; INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:

```

```

;
LENGTH: 3560 base pairs

```

```
STRANDEDNESS: double
PROTOCOLV: 140000
```

MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1953
 US-08-121-713D-59

alignment_scores:
 Quality: 59.00 Length: 104
 Ratio: 1.035 Gaps: 6
 Percent Similarity: 54.808 Percent Identity: 26.923

alignment_block:

US-10-048-197-2 x US-08-121-713D-59/rev ..

Align seg 1/1 to reverse of: US-08-121-713D-59 from: 1 to: 3560

```

6  TlAargLeuThrIleSerAlaLeuThrAlaLeuValThrGlyC 22
   :::::  :::::  :::::  :::::  :::::  :::::
1650 CTCGGTATCCGATACGACAGATTATCATCTCTGTGACA..... 1606

22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnThrIleG 39
   :::::  :::::  :::::  :::::  :::::  :::::
1605 .ACGGCAGCCGAGAGATAGCGCTGTAAAGAACCCGACCAAGGAAAG 1557

39 LglInThrIleIleLysGlyThrAsnLysGlnIu.IleSerSerAr 55
   :::::  :::::  :::::  :::::  :::::  :::::
1556 ATCGAAC.....GGCCAGAACGCCATCAGCGAGGCTCTCCACAGT 1516

55 gpheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG 72
   :::::  :::::  :::::  :::::  :::::  :::::
1515 GTAC..... 1512

72 LylsThrAlaIleLeuAlaProAsnArgTrpGlnIuIleLeuSerLeu 88
   :::::  :::::  :::::  :::::  :::::  :::::
1511 .....TGGGGTGTATATCTTCAAGTCTTATCAATATGTGATT 1467

89 IleIle...SerPheLeuTrpValLysProTyrArgProLysAsnLeu 104
   :::::  :::::  :::::  :::::  :::::  :::::
1466 ATTTCCTGATCTGTCTGTG.....CGTCGGCATTCATTAAGTC 1426

104 rPheTyrLeu 107
   :::::  :::::  :::::  :::::  :::::  :::::
1425 CATGTCTGTG 1416
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-835-268-59

seq_documentation_block:

; Sequence 59, Application US/08835268
 ; Patent No. 5807826

; GENERAL INFORMATION:

; APPLICANT: Goodman, Corey S.
 ; APPLICANT: Kolodkin, Alex L.
 ; APPLICANT: Mathes, David R.
 ; APPLICANT: Bentley, David R.
 ; APPLICANT: O'Connor, Timothy
 ; TITLE OF INVENTION: The Semaphorin Gene Family
 ; NUMBER OF SEQUENCES: 100
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
 ; STREET: 268 Bush Street, Suite 3200
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/835,268
 FILING DATE:

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/121,713
 FILING DATE: 13-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A.
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: B94-002-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415)343-4341
 TELEFAX: (415) 343-4342
 TELEX:
 INFORMATION FOR SEQ ID NO: 59:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3560 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1953
 US-08-835-268-59

alignment_scores:
 Quality: 59.00 Length: 104
 Ratio: 1.035 Gaps: 6
 Percent Similarity: 54.808 Percent Identity: 26.923

alignment_block:

US-10-048-197-2 x US-08-835-268-59/rev ..

Align seg 1/1 to reverse of: US-08-835-268-59 from: 1 to: 3560

```

6  TlAargLeuThrIleSerAlaLeuThrAlaLeuValThrGlyC 22
   :::::  :::::  :::::  :::::  :::::  :::::
1650 CTCGGTATCCGATACGACAGATTATCATCTCTGTGACA..... 1606

22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnThrIleG 39
   :::::  :::::  :::::  :::::  :::::  :::::
1605 .ACGGCAGCCGAGAGATAGCGCTGTAAAGAACCCGACCAAGGAAAG 1557

39 LglInThrIleIleLysGlyThrAsnLysGlnIu.IleSerSerAr 55
   :::::  :::::  :::::  :::::  :::::  :::::
1556 ATCGAAC.....GGCCAGAACGCCATCAGCGAGGCTCTCCACAGT 1516

55 gpheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG 72
   :::::  :::::  :::::  :::::  :::::  :::::
1515 GTAC..... 1512

72 LylsThrAlaIleLeuAlaProAsnArgTrpGlnIuIleLeuSerLeu 88
   :::::  :::::  :::::  :::::  :::::  :::::
1511 .....TGGGGTGTATATCTTCAAGTCTTATCAATATGTGATT 1467

89 IleIle...SerPheLeuTrpValLysProTyrArgProLysAsnLeu 104
   :::::  :::::  :::::  :::::  :::::  :::::
1466 ATTTCCTGATCTGTCTGTG.....CGTCGGCATTCATTAAGTC 1426

104 rPheTyrLeu 107
   :::::  :::::  :::::  :::::  :::::  :::::
1425 CATGTCTGTG 1416
```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-060-692-59

seq_documentation_block:

; Sequence 59, Application US/09060692
 ; Patent No. 5935865

; GENERAL INFORMATION:

; APPLICANT: Goodman, Corey S.
 ; APPLICANT: Kolodkin, Alex L.
 ; APPLICANT: Mathes, David
 ; APPLICANT: Bentley, David R.

```

APPLICANT: O'Connor, Timothy
TITLE OF INVENTION: The Semaphorin Gene Family
NUMBER OF SEQUENCES: 100
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/060,692
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/121,713
FILING DATE: 13-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B94-002-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415) 343-4342
TELEX:
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 3560 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1953
US-09-060-692-59

```

```

alignment_scores:
Quality: 59.00 Length: 104
Ratio: 1.035 Gaps: 6
Percent Similarity: 54.808 Percent Identity: 26.923

```

alignment_block:

US-10-048-197-2 x US-09-060-692-59/rev ..

Align seg 1/1 to reverse of: US-09-060-692-59 from: 1 to: 3560

```

6 llaargleuthrlleSeraAlaLeuThrAlaLeuValThrglcy 22
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1650 CTCGCGATCCGATCGACGAGATTATCATCTCTCTGTGACA..... 1606
22 svalsertthnglyasnvalalemetlysglugainglthrlleg 39
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1605 .ACGGCGACCGCAGAGTACCTGTAAAGACCGACGACGAGAAAG 1557
39 luginthrllelelysglylysthrasnlysglgu..lleseratar 55
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1556 ATCGAAC.....GCCGACAGCGCCATCACGCGCTCTCCACAGT 1516
55 gpheglyserAlaAspSerlleserphemetIleValIlelyspheg 72
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1515 GTAC..... 1512
72 lythshthrlleleleAlaProAsnArgrpGlnglulileuSerleu 88
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1511 .....TGGCGTTGATATATCTTCAAAAGTTCTTATCAATATTGCGATT 1467

```

```

89 llele...serpheleuthrVallyspRofyArGProlysasnleuSe 104
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1466 ATTTCGTGATCTCTGCTCTG.....CGTCGCAATCAATAATGTC 1426
104 rpehlyrleu 107
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1425 CATGCTGCTG 1416

```

seq_name: /cgn2_6/plodata/2/lna/6A_COMB.seq:US-08-833-391-59

seq_documentation_block:

```

Sequence 59, Application US/08833391
Patent No. 6013781
GENERAL INFORMATION:
APPLICANT: Goodman, Corey S.
APPLICANT: Kolodkin, Alex L.
APPLICANT: Mathes, David
APPLICANT: Bentley, David R.
APPLICANT: O'Connor, Timothy
TITLE OF INVENTION: The Semaphorin Gene Family
NUMBER OF SEQUENCES: 100
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/833,391
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/121,713
FILING DATE: 13-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B94-002-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415) 343-4342
TELEX:
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 3560 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1953
US-08-833-391-59

```

```

alignment_scores:
Quality: 59.00 Length: 104
Ratio: 1.035 Gaps: 6
Percent Similarity: 54.808 Percent Identity: 26.923

```

alignment_block:

US-10-048-197-2 x US-08-833-391-59/rev ..

Align seg 1/1 to reverse of: US-08-833-391-59 from: 1 to: 3560

```

6 llaargleuthrlleSeraAlaLeuThrAlaLeuValThrglcy 22
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1650 CTCGGATGCGGATAGCGGACAGATTATCATCTGCTCTTGAGAC.... 1606
22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleG 39
      ::::|||||:|||||:|||||:|||||:|||||:
1605 .ACGGGACCGCGGAAAGTATCCGTAAAGACCGACACACGGAAG 1557
39 IuGlnThrIleIleLysGlyLysThrAsnLysGlnGlu.IleSerSer 55
      :|||:|||||:|||||:|||||:|||||:
1556 ATCGAACC.....GCCCAACCGGCATCAGAGGCTCTCCACAGT 1516
55 gpheGlySerAlaAspSerIleSerPheMetIleValIleLysPheG 72
      ::
1515 GTAC..... 1512
72 LysIleThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeu 88
      :|||:|||||:|||||:|||||:|||||:
1511 ....TGGCGGTGATTAATCTCAAGGTCTTAATAAATGTCGATTT 1467
89 IleIle...SerPheLeuTrpValLysProTyrArgProLysAsnLeu 104
      |||:|||||:|||||:|||||:|||||:
1466 ATTTCCTGATCTCTGCTCTG.....CGTCGCAATTCATTAATGTC 1426
104 rPheTyrLeu 107
      ::::||||
1425 CARGTCGTTG 1416
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-060-610-59

```

```

seq documentation block:
: Sequence 59, Application: us/09060610
: Patent No. 6344544
:
: GENERAL INFORMATION:
: APPLICANT: Goodman, Corey S.
: APPLICANT: Koldackin, Alex L.
: APPLICANT: Mattes, David
: APPLICANT: Bentley, David R.
: APPLICANT: O'Connor, Timothy
: TITLE OF INVENTION: The Semaphorin Gene Family
:
: NUMBER OF SEQUENCES: 100
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 268 Bush Street, Suite 3200
: CITY: San Francisco
: STATE: CA
:
: COUNTRY: USA
: ZIP: 94104
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/060,610
: FILING DATE:
:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/835,268
: FILING DATE:
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman, Richard A.
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: B94-002-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415)343-4341
: TELEFAX: (415) 343-4342
:
: TELEX:
:
: INFORMATION FOR SEQ ID NO: 59:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3560 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
: MOLECULE TYPE: cDNA
:

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1953
US-09-060-610-59

```

```
alignment_scores:
  Quality: 59.00      Length: 104
  Ratio: 1.035      Gaps: 6
Percent Similarity: 54.808      Percent Identity: 26.923
```

```
alignment_block:
  pc-10-048-107-3 v pc-09-060-610-59/rev
```

Align seg 1/1 to reverse of: US-09-060-610-59 from: 1 to: 3560

6 ILeArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCys 22

1650 CTCGGTATCCGGATACGGCAGATTATCATCTCTGTCCTTGTGACA..... 1606

22 svaIserThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleG 39

1605 .ACGGCGACCGCAGAGTAGCCTGTAAAGAGGCCGACCAGCAGCGGAAAG 1557

39 lucInThrlelelysglysthrasnlysgIngu. lIeserSerAr 55

1556 ATCGAAC.....GGCCAGAACGGCCATCAGAGGGTCTCCACAGT 1516

55 gpheglyserAlaaspSerIleSerPheMetIleValIleIleLysPheG 72

1515 GTAC..... 1514

72 L Y N I S T H R A I A I L E U A L P R O A S N A R G T T P G I N G I U I L E U S E R L E U 88
|||||.....

1511TGGCGTTGATTATATCTCAAGTTCCTATCAATATTGTGATT 1467

```

89 IleIle...SerPheLeuTrpValLysProTyrArgProLysAsnLeuSe 104
|||...|||:::|

```

1466 ATTTCCTGATCCTTGCTCTGG.....CGTCGCGAATCCAATAAGTC 1426

104 rphETyrlEu 107
|::|||

1425 CATGTCGTTG 1416

seq_name: /cgn2_6/ptodata/2/lna/PCTUS_COMB.seq:PCT-US94-10151A-3

```
seq_documentation_block:
: Sequence 59. Application PC/TTTS9410151A
```

GENERAL INFORMATION:
APPLICANT: The Regents of the University of California

```

; TITLE OF INVENTION: The Semaphorin Gene Family
;
; NUMBER OF SEQUENCES: 66

```

ADDRESSEE: FLEHR HOHBACH TEST ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco

```
STATE: CA
COUNTRY: USA
```

ZIP: 94111-4187

```

;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10151A
;

```

FILED DATE: 13-SEP-1994
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.

REGISTRATION NUMBER: 36,62/
REFERENCE/DOCKET NUMBER: FP-58750-PC/RAO

TELECOMMUNICATION INFORMATION:

TYPE: DNA

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-210-361-1

LOCATION: (9)...(4434)

alignment_block

US-10-048-197-2 x US-09-210-361-1 .

OM of: us-10-048-197-2 to: N.Geneseq_032802: * out_format : pfs
 Date: Sep 18, 2002 5:51 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
 -MODEL=frame+g2n.model -DEV=xlip
 -O=/g9n2.1/USPTO.spool/US10048197/runat_17092002.135850.22677/app-query.fasta.1.168
 -DB=N.Geneseq_032802 -QFMT=fastap -SUFFIX=std.rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
 -OGAPEXT=0.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPEXT=10.000 -YGAPEXT=0.500
 -DELDP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blissum62
 -TRANS=human40.cdt -LIST=45 -DOCALLION=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USER=US10048197 -CGN1=1.523 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLIPX -WAIT -THREADS=1

Search information block:
 Query: US-10-048-197-2
 Query length: 111
 Database: N.Geneseq_032802: *
 Database sequences: 1736436
 Database length: 858457221
 Search time (sec): 278.560000

Sequence	Strd Orig	ZScore	EScore len	Documentation
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF59778 +		558.00	1386.78	6.8e-69
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF28551 +		558.00	1323.03	2.4e-65
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV03597 +		80.50	188.55	0.0373
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV15394 +		80.50	188.55	0.0373
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV15394 +		80.00	184.60	0.0619
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV15394 +		80.00	184.60	0.0619
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV15394 +		78.00	188.53	0.4861
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV25208 +		74.50	143.66	11.82
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF28550 +		71.00	147.58	7.14
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:ABL10148 +		70.00	134.93	36.19
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL12634 +		70.00	133.67	42.54
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV13387 +		70.00	130.90	60.42
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF28552 +		69.50	153.51	3.34
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF28552 +		69.50	153.44	3.37
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF28552 +		69.50	147.57	7.15
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV21612 +		69.00	101.30	2.7e+03
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV21612 +		69.00	78.52	5.0e+04
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV21612 +		68.00	68.66	1.7e+05
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAV03041 +		68.00	53.19	4.7e+06
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV03041 +		68.00	150.73	4.77
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV03041 +		68.00	148.27	6.54
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV03041 +		68.00	140.36	18.02
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV21310 +		68.00	133.58	43.04
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV21310 +		68.00	132.77	47.74
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV22205 +		68.00	111.67	712.69
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV22205 +		67.50	140.33	18.09
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV22205 +		67.50	135.95	31.75
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV22205 +		67.50	132.55	49.07
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV22205 +		67.50	126.87	101.78
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV19980 +		67.50	124.78	132.94
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV211169 +		67.50	124.78	132.94
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV211214 +		67.50	124.78	132.94
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAF31912 +		67.50	111.75	707.51
/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAF31912 +		67.50	111.75	707.51

/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:ABL33364 - 67.50 108.34 1.1e+03
 /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV21612 + 67.00 134.79 36.85
 /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV14505 + 67.00 129.96 68.41
 /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV14505 + 67.00 129.53 72.35
 /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL11270 + 67.00 128.84 79.06

seq_name: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF59778

seq_documentation_block:

ID AAF59778 standard; DNA: 336 BP.

AAF59778:

02-MAY-2001 (first entry)

Moraxella catarrhalis strain ATCC43617 BASB122 DNA.

BASB122 protein; strain ATCC43617; antigen; antibody; vaccine;

genetic immunisation; infection; upper respiratory tract; otitis media;

hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;

invasive disease; antibacterial; auditory; ds.

Moraxella catarrhalis.

MO200109337-A2.

08-FEB-2001.

31-JUL-2000; 2000MO-EP07365.

30-JUL-1999; 99GB-0018034.

30-JUL-1999; 99GB-0018036.

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Thonard J;

WPI: 2001-159874/16.

P-P-SDB; AAB60640.

New BASB122 and BASB124 polypeptides and polynucleotides from Moraxella

catarrhalis strain ATCC 43617, useful as therapeutic agents or vaccines

against bacterial infections, e.g. otitis media or pneumonia

Claim 13: Page 66; 75pp; English.

The invention relates to the Moraxella catarrhalis strain ATCC43617

BASB122 and BASB124 proteins (AAB60640 and AAB60641, respectively)

and to DNA encoding them (AAF59778 and AAF59779, respectively). The

invention also relates to immunogenic fragments of the BASB122 and

BASB124 proteins, expression vectors and host cells comprising BASB122

or BASB124 nucleic acids, the recombinant production of BASB122 or

BASB124, vaccine compositions comprising the BASB122 or BASB124 proteins

or nucleic acids, an antibody against BASB122 or BASB124, therapeutic

compositions comprising an anti-BASB122 or anti-BASB124 antibody, and a

method of identifying a Moraxella catarrhalis infection via the

detection of BASB122 or BASB124 proteins or antibodies. The vaccine

compositions of the invention are useful as prophylactic or therapeutic

agents against Moraxella catarrhalis infections in mammals, particularly

humans. Moraxella catarrhalis is a Gram negative bacterium frequently

isolated from the human upper respiratory tract, which is responsible for

several pathological conditions. It is responsible for about 15% of

otitis media cases in children (which can lead to temporary or permanent

hearing loss). It also causes pneumonia in elderly people, and sinusitis,

nosocomial infections and, less frequently, invasive diseases. BASB122 or

BASB124 proteins or nucleotides may additionally be used in screening for

novel antibacterial compounds, and in the diagnosis and staging of

infections. The present sequence represents DNA encoding the Moraxella

catarrhalis strain ATCC43617 BASB122 protein.

Sequence 336 BP; 106 A; 65 C; 61 G; 104 T; 0 other;

alignment_scores: Quality: 558.00 Length: 111
 Ratio: 5.027 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AAF59778 ..

Align seg 1/1 to: AAF59778 from: 1 to: 336

1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
1 ATGTTAAATCATCATATATTCGCTGACATATTCGCTTACTGACGGCAGT 50
17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGln 34
1 TTTGGTAACAGGTTGTTCTTCTACTGTAATGTTGCAATGAAGAGCAA 100
34 sngInGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
101 ACCAACAAACCATGACAGACACCATCATTAAGGCAAGACCATTAAGCAA 150
51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVa 67
151 GAGATTTCTGATGATTTGGTCTGCTGATAGCATCTCTTTATGATAGT 200
67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 84
201 GGTAAATTAATTTGGACATACCGCATACCTGCTCCAAACCGATGGCAAG 250
84 lIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
251 AAATTTATCCCTATATATTTCTTTCTTTGGTGAACCATACAGACCA 300
101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
301 AAGAACTTGTCATTTTATTTTGACAGCAAAAGCA 333

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF28551

seq_documentation_block:
ID AAF28551 standard; DNA; 94750 BP.
XX AAF28551;
AC AAF28551;
XX
XX 04-APR-2001 (first entry)
XX
XX Genomic fragment #38.
XX
XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
XX
XX Moraxella catarrhalis.
OS
XX
XX WO200078968-A2.
XX
XX
XX 28-DEC-2000.
XX
XX 16-JUN-2000; 2000MO-US16649.
XX
XX 18-JUN-1999; 9905-0140121.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Lagace RE, Patterson C, Berg KL;
XX
XX WPI; 2001-041427/05.
XX
XX Genomic library for identifying diagnostic and therapeutic
XX compositions, and for identifying virulence factors, regulatory
XX elements and drug targets, comprises Moraxella catarrhalis nucleic
XX acids -
XX

PS Claim 1; Page 415-436; 545pp; English.

XX
XX The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
XX

Sequence 94750 BP; 27128 A; 18811 C; 21444 G; 27367 T; 0 other;

alignment_scores: Quality: 558.00 Length: 111
 Ratio: 5.027 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AAF28551 ..

Align seg 1/1 to: AAF28551 from: 1 to: 94750

1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
75267 ATGTTAAATCATCATATTCGCTGACATATTCGCTTACTGACGGCAGT 75316
17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGln 34
75317 TTTGGTAACAGGTTGTTCTTCTACTGTAATGTTGCAATGAAGAGCAA 75366
34 sngInGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
75367 ACCAACAAACCATGACAGACACCATCATTAAGGCAAGACCATTAAGCAA 75416
51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVa 67
75417 GAGATTTCTGATGATTTGGTCTGCTGATAGCATCTCTTTATGATAGT 75466
67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 84
75467 GGTAAATTAATTTGGACATACCGCATACCTGCTCCAAACCGATGGCAAG 75516
84 lIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
75517 AAATTTATCCCTATATATTTCTTTCTTTGGTGAACCATACAGACCA 75566
101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
75567 AAGAACTTGTCATTTTATTTTGACAGCAAAAGCA 75599

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV03597

seq_documentation_block:
ID AAV03597 standard; DNA; 267 BP.
XX AAV03597;
AC AAV03597;
XX
XX 22-OCT-1998 (first entry)
XX
XX Neisseria meningitidis DNA sequence E29.
XX
XX N. gonorrhoeae; N. lactamica; chromosome Z2491; region 1; region 2;
KW region 3; pathogenicity; blood-brain barrier; diagnosis; infection;
XX meningitis; ss.
XX
XX Neisseria meningitidis.
OS
XX
XX WO9802547-A2.
XX
XX

73 sThrAlaIleLeu 77
 ::|||::|||
 211 CAATGCTATGTGA 223

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA76029

seq_documentation_block:
 ID AAA76029 standard; DNA: 339 BP.

XX AAA76029;

DT 19-JAN-2001 (first entry)

DE Neisseria meningitidis BASB059 gene.

KM BASB059: respiratory tract infection; invasive bacterial disease;

KW bacteraemia; meningitis; ds.

OS Neisseria meningitidis.

XX Key Location/Qualifiers

FT CDS 1..339

FT /product= "BASB059"

XX WO200044904-A1.

PD 03-AUG-2000.

PF 25-JAN-2000; 2000WO-EP00561.

PR 29-JAN-1999; 99GB-0002070.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Thonnard J;

XX WPI; 2000-505978/45.

DR P-PSDB; AAB21107.

XX New isolated polypeptide from Neisseria meningitidis is useful for

PT detection and treatment of N. meningitidis infection -

XX Claim 7: Page 61: 77pp; English.

XX The present sequence is the Neisseria meningitidis BASB059 gene. The

CC gene, its protein, agonists and antagonists can be used in diagnostic

CC assays for the bacterium, as vaccines to prevent infection and as

CC treatments for bacterial infection, particularly those caused by

CC Neisseria meningitidis, such as upper respiratory tract infections,

CC invasive bacterial diseases, bacteraemia and meningitis.

XX Sequence 339 BP: 93 A; 84 C; 84 G; 78 T; 0 other;

SO

Align seg 1/1 to: AAA76029 from: 1 to: 339

7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuValThrGlyCysVa 23

10 CGCATTCATCGTTTCGGCTGCGTTCGTTGCGATTCAGCAGGTTCGGC 59

23 ISeThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGlu 40

60 CTCATCATCAATATGTAACCGTTTCGACCGAACAACCTTCAGGAAGTCGCG 109

alignment_scores: Quality: 80.00 Length: 89

Percent Similarity: 61.798 Percent Identity: 29.213

alignment_block:

US-10-048-197-2 x AAA76029 ..

Align seg 1/1 to: AAA76029 from: 1 to: 339

7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuValThrGlyCysVa 23

10 CGCATTCATCGTTTCGGCTGCGTTCGTTGCGATTCAGCAGGTTCGGC 59

23 ISeThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGlu 40

60 CTCATCATCAATATGTAACCGTTTCGACCGAACAACCTTCAGGAAGTCGCG 109

40 InThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
 ::|||::|||
 110 CGTTTGCCCTTGCGGCTCAGCAAAATGCCGTAATAAATCAGCAACCGCAGC 159
 57 GlySerAlaAspSerIleSerPheMetIleValIleLysPheGlyH 73
 ::|||::|||
 160 AATGAACCATACGATCAGCACTTACCGCATGTGGGTAGACCGCGAG 209
 73 IStThrAlaIleLeuAlaProAsnArgTrpGlnGluIleSerLeuIle 89
 ::|||::|||
 210 CCAATGCTATGTATCCAGCTGTAATCAGCAATCGCGCTTACCACTTCGC 259
 90 IleSerPheLeuTrp 94
 ::|||::|||
 260 ATGCATTTGTTTGG 274

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA15300

seq_documentation_block:
 ID AAA15300 standard; DNA: 339 BP.

XX AAA15300;

DT 04-SEP-2000 (first entry)

DE DNA encoding a polypeptide of a Neisseria pathogenic strain.

KW pathogenic strain; Neisseria; vaccine; Neisseria infection; ss.

OS Neisseria meningitidis.

XX Key Location/Qualifiers

FT CDS 1..339

FT /product= a

XX WO200026375-A2.

PD 11-MAY-2000.

PF 28-OCT-1999; 99WO-FR02643.

PR 30-OCT-1998; 98FR-0013693.

PA (INRM) PASTEUR MERIEUX SERUMS & VACCINS SA.

PI (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aujaune L, Bouchardon A, Renauld-Mongenie G, Rokhl B, Nassif X;

XX Tinsley C, Perrin A;

XX WPI; 2000-365622/31.

DR P-PSDB; AAY93269.

XX New polypeptide specific for pathogenic Neisseria useful in therapeutic

PT or preventative vaccines and for diagnosis -

XX Claim 2: Page 69: 187pp; French.

XX The present sequence encodes a protein that is specific for pathogenic

CC strains of Neisseria. The polynucleotides, polypeptides, or their

CC antigenic fragments, are used in vaccines to treat or protect against

CC Neisseria infections, particularly by N. meningitidis. The

CC polynucleotide sequence is also used for recombinant production of

CC the polypeptide and to produce attenuated Neisseria strains that

CC overexpress it, or express it in a non-toxic mutant form.

SO Sequence 339 BP: 93 A; 84 C; 84 G; 78 T; 0 other;

alignment_scores: Quality: 80.00 Length: 89

Percent Similarity: 61.798 Percent Identity: 29.213

alignment_block:

US-10-048-197-2 x AAA15300 ..

Align seg 1/1 to: AAA15300 from: 1 to: 339

```

7 ArgLeuThrIleSerAlaLeuThrAlaLeuLeuValThrGlyCysVa 23
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
10 CGCATCATGCTTCCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 59
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
23 lSerThrGlyAsnValAlaMetIysGlnGlnAsnGlnGlnThrIleGln 40
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
60 CTCATCAATCAATTAATGTAACCTTCCGACCAAGAACTTACAGAACGTC 109
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
40 lInThrIleIleGlyGlyThrAsnIysGlnGlnIleSerSerArgPhe 56
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
110 CGTTTCCCTGGGCGTCAGCCAAATGCCGTAATAATACACCAACCCGAGC 159
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
57 GlySerAlaAspSerIleSerPheMetIleValIleIysPheGlyH 73
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
160 AATGAAGATACGATACCATCACTTACCCGCAACTGTGGGTAGCGCGTAG 209
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
73 lSerThrAlaIleLeuAlaProAsnArgTrpGlnGlnIleLeuSerIle 89
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
210 CCAATGCTATGTTACGAGTGTATATCAGACAAATCGCGCTTACCATTC 259
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
90 lIleSerPheLeuTrp 94
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
260 ATGCAATTTGTTTGG 274
  |||:::|||||:::|||||:::|||||:::|||||:::|||||

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA05521

seq_documentation_block:

ID AAA05521 standard; DNA; 900 BP.

```

XX AC AAA05521;
XX DT 24-MAY-2000 (first entry)
XX DE Streptococcus pneumoniae type 4 nucleotide sequence 4170.6.
XX KM Streptococcus pneumoniae; vaccine; screening; protein antigen;
XX KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
XX KM pneumococcal disease; ds.
XX OS Streptococcus pneumoniae.
XX PN WO200006737-A2.
XX PD 10-FEB-2000.
XX PF 27-JUL-1999; 99WO-GB02451.
XX PR 27-JUL-1998; 98GB-0016337.
XX PR 19-MAR-1999; 99US-0125164.
XX PA (MICR-) MICROBIAL TECHNIQS LTD.
XX PI Gilbert CFE, Hansbro PM;
XX DR WPI, 2000-195300/17.
XX PT New Streptococcal protein, useful as a vaccine, for diagnosis of
XX PT pneumococcal diseases and for screening agents capable of antagonizing
XX PT or inhibiting expression of the protein
XX PS Claim 7; Page 54; 108pp; English.
XX CC AAV81501 to AAV81679 represent specifically claimed protein sequences
XX CC isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
XX CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
XX CC The sequences have antibacterial and antiinflammatory properties.
XX CC The protein sequences, and fragments of them, are useful as immunogens

```

CC and/or antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also
 CC useful for screening an agent capable of antagonizing, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
 CC and meningitis. AAA05591 to AAA05614 represent primers used in the
 CC exemplification of the present invention.

SQ Sequence 900 BP; 240 A; 155 C; 211 G; 294 T; 0 other;

alignment_scores:

	Quality:	78.00	Length:	127
	Ratio:	1.200	Gaps:	6
	Percent Similarity:	51.181	Percent Identity:	28.346

alignment_block:

US-10-048-197-2 x AAA05521 ..

Align seg 1/1 to: AAA05521 from: 1 to: 900

```

10 lIleSerAlaLeuThrAla...LeuLeuValThrGlyCysValSerTh 25
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
232 CTAAGAGCCTTGGTGAACAGCTGTGATCTCTGAACGGGCTGTCTAGT 281
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
25 r.....GlyAsnValAlaMetIysGlnGlnAsnGlnGlnThrIleGln 40
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
282 CATTGGAAATGTCACGAAGATTTTGCATCCGCAACGATCAATGATG 331
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
40 lInThrIleIle.....
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
332 AGGGGATTCCTCTGTAGGAATATTGCGATTAATCAATCTGTATGCG 381
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
44 .....LysGlyIysThrAsnIysGlnGlnIleSerSerArg 55
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
382 AGTCTGCTGCTGTGTAAGGGAAGCAAGAAATGATGCTATCTGATCT 431
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
55 gPheGlySerAlaAspSerIleSerPhe.....MetIleV 67
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
432 GCATTTTCTGGAGATACGCTAGGCTGTGTAAGCTTTATCTGTAGCGA 481
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
67 aValAlaIlePheGlyIleThrAlaIleLeuAlaProAsnArgTrpGln 83
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
482 TTGCTTCATTTACGAGACTGGTATCTCTAGATCT..... 519
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
84 GluIleLeuSerIleIleSerPheLeuTrpValIysProTyraArgPr 100
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
520 ...CTTTTGTCCCTTGTGATTTCTTTTATTTCTTCAAAAGCCCTTCC 566
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
100 GlyAsnIleSerPheTrpLeuThrAlaIys 110
  |||:::|||||:::|||||:::|||||:::|||||:::|||||
567 ACCT.....TTTGTGCTACACTCAAG 588
  |||:::|||||:::|||||:::|||||:::|||||:::|||||

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV52208

seq_documentation_block:

ID AAV52208 standard; DNA; 8136 BP.

```

XX AC AAV52208;
XX DT 23-OCT-1998 (first entry)
XX DE Streptococcus pneumoniae genome fragment SEQ ID NO:75.
XX KM Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
XX KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX OS Streptococcus pneumoniae.
XX PN WO9818931-A2.
XX PD 07-MAY-1998.

```

```
XX 30-OCT-1997; 97WO-US19588.
PF
XX 31-OCT-1996; 96US-0029960.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;
XX WPI: 1998-272225/24.
DR
XX Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae

Claim 1; Page 617-622; 1409pp; English.

The present invention describes a computer readable medium which has
the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
recorded on it, or a representative fragment or a sequence at least 95%
identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
Streptococcus pneumoniae. The present invention also describes an
isolated nucleic acid molecule encoding a homologue of any of the
fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
nucleic acid molecule is produced by a process comprising: (a) screening
a genomic DNA library using as a probe a target sequence defined by any
of the sequences in SEQ ID NO:1 to 391, identifying members of the
library which contain sequences that hybridise to the target sequence and
isolating the nucleic acid molecules from the members; or (b) isolating
mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
molecules whose nucleotide sequence is homologous to amplification
primers derived from the fragment of the S. pneumoniae genome to prime
the amplification and isolating the amplified sequences. The computer
readable medium can be used in a computer-based system for identifying
fragments of the S. pneumoniae genome of commercial importance, or
expression modulating fragments of the S. pneumoniae genome. Products
from the present invention can be used in diagnosis kits and assays, and
pharmaceutical compositions and vaccines for S. pneumoniae.

Sequence 8136 BP; 2249 A; 1481 C; 1983 G; 2423 T; 0 other:

alignment_scores:
Quality: 78.00 Length: 127
Ratio: 1.200 Gaps: 6
Percent Similarity: 51.181 Percent Identity: 28.346

alignment_block:
US-10-048-197-2 x AAV52208 ..

Align seg 1/1 to: AAV52208 from: 1 to: 8136

10 lIeSerAlaLeuLeuThraAla...LeuLeuValThrGlyCysValSerTh 25
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1926 CTAAGAGCGCTTGCTGATACCTGATTCGTAACGGCTGCTGTTCTACT 1975
25 r.....GlyAsnValAlaMetLysGluLAsnGlnGlnThrIleGln 40
: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
1976 CATTTCGGAATAATGTCACGAGATTTTCATCCGCAACGATCAATGATG 2025
40 lntThrIlele..... 43
:::|||||:::
2026 ACGGCAATTCCTGCTGATTAGCAATTATTCGATCACTCAATCTGTTAGCG 2075
44 .....LysGlyLysThrAsnLysGlnGlnLysSerSer 55
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
2076 AGTCGTGCTGTTGTAAGCAAGCAACAAATGATGATCTATCTGAGCTCT 2125
55 gpheGlySerAlaAspSerIleSerPhe.....MetIleVal 67
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

```
2126 GCATTTCTGGAGATACGCTAGGTGGTGGTATTCCTGATGCGCA 2175
67 aValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 83
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
2176 TTTCTTCGATTTACGACTGCTATATCTAGATCTCT..... 2213
84 GluIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPr 100
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
2214 ...CTTTGTCCTGCTGATTTCTTTCTTATCTTCAAAAGCCCTTCC 2260
100 oLysAsnLeuSerPheTyrLeuThrAlaLys 110
::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
2261 ACCT.....TTTGGTCACTACCTCAAG 2282

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AA562620

seq_documentation_block:
ID AAS62620 standard; cDNA; 1709 BP.
AC AAS62620;
XX
XX 14-FEB-2002 (first entry)
DE
XX cDNA sequence #407 encoding novel human secreted protein.
KW
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;
KW immunosuppressive; antineumatic; ss.
OS
XX Homo sapiens.
XX
XX WO200177291-A2.
XX
XX 18-OCT-2001.
XX
XX 29-MAR-2001; 2001WO-US10485.
XX
XX 06-APR-2000; 2000US-195604P.
XX
XX (GENY ) GENETICS INST INC.
XX
XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
XX Gulikota K, Graham JR;
XX
XX WPI: 2002-010900/01.
XX
XX New polynucleotides encoding secreted proteins useful for treating e.g.
XX asthma, HIV and Crohn's disease -
XX
XX Claim 1; Page 289-290; 391pp; English.
XX
XX The present invention relates to the isolation of novel cDNA sequences
XX which encode human secreted proteins. The cDNA sequences have been
XX derived from a variety of human tissues. The invention also provides
XX a method for producing proteins from these polynucleotide sequences.
XX The proteins are useful for identifying compounds that modulate their
XX activity and production, and the cell is also useful for identifying
XX compounds that modulate expression of the polynucleotide sequences
XX encoding the secreted proteins. The sequences of the invention are
XX useful for treating diseases such as hyperproliferative disorders
XX (e.g. cancer), immune deficiency disorders (e.g. severe combined
XX immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
XX sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory
XX disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis).
XX The polynucleotide sequences of the invention are also useful in gene
XX therapy. AAS62214-AAS62838 represent the cDNA sequences of the
XX invention that encode for novel human secreted proteins.
XX
XX Sequence 1709 BP; 481 A; 358 C; 421 G; 449 T; 0 other:

alignment_scores:

```

```

PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PR acids -
XX
XX Claim 1; Page 391-415; 545bp; English.
XX
CC The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
XX Sequence 99629 BP; 29233 A; 19222 C; 21909 G; 29264 T; 1 other:
XX
XX alignment_scores:
XX      Quality: 71.00      Length: 109
XX      Ratio: 1.224      Gaps: 4
XX      Percent Similarity: 53.211      Percent Identity: 22.936
XX
XX alignment_block:
XX US-10-048-197-2 x AAF28550/rev ..
XX
XX Align seg 1/1 to reverse of: AAF28550 from: 1 to: 99629
XX
XX 4 HISHSHIleArgLeuThrIleSerAlaIleuThrAlaIleuValrh 20
XX 76708 TATCAAAATTTCCAGCAGCCCTGAACCTCACTCTTGAGGACGTGCATGAT 76659
XX      ::::::::::::::::::::
XX 20 rclYcysValSerThrGlyAsnValAlaMetLysGluGlnGlnGln 37
XX 76658 AAGCGTCTGTGATACACATATGCTGTGGTGTACCAAAATTCATC 76609
XX      ::::::::::::::::::::
XX 37 htrIleGluIleThrIle.....IleLysGlyLysThrAsnLysGlnGlu 51
XX      ::::::::::::::::::::
XX 76608 AAGTTCATACACCCACTGAGATGATGAACCTTAACGCTTCATGAACAG 76559
XX      ::::::::::::::::::::
XX 52 IleserSerArgPheGlySerAlaIleuValSerIleSerhemeIleValVa 68
XX 76558 GCAAAAACCTTGGCGTGC..... 76541
XX      ::::::::::::::::::::
XX 68 IleLysPheGlyHisThrAlaIleuAlaProAsnArgTrpGlnGln 85
XX      ::::::::::::::::::::
XX 76540 .....TGGTATCCCAATATGCGATGCCAAATTCGACAGCGC 76501
XX      ::::::::::::::::::::
XX 85 IeLeu.....SerLeuIleIleSerPheLeuTrpValLysProTyr 98
XX      ::::::::::::::::::::
XX 76500 TCGTAAAGCTACGCCCACTGATACCGCACGACATGATACGCCCA... 76454
XX      ::::::::::::::::::::
XX 99 ArgProLysAsnLeuSerPheTyrLeu 107
XX      ::::::::::::::::::::
XX 76453 AAGCCATAGAAATTAATCCAATTTT 76427
XX
XX seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL10149
XX
XX seq_documentation_block:
XX ID ABL10149 standard; cDNA; 969 BP.
XX
XX AC ABL10149;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster expressed polynucleotide seq ID NO 24929.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX OS Drosophila melanogaster.
XX
XX

```

PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR P-PSDB; ABB66046.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS
 PS Claim 1; SEQ ID NO 24929; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
 CC sequences (AB57737-AB572072), and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 CC Sequence 969 BP; 271 A; 228 C; 249 G; 221 T; 0 other;
 SQ
 alignment_scores:
 Quality: 70.00 Length: 41
 Ratio: 2.500 Gaps: 2
 Percent Similarity: 68.293 Percent Identity: 43.902
 alignment_block:
 US-10-048-197-2 x ABL10149/rev ..
 Align seg 1/1 to reverse of: ABL10149 from: 1 to: 969
 54 SerArpPheGlySerAlaAspSerIleSerPheMetIleValIlely 70
 ||||||| :: |||:|||||: |||
 882 AGTCGTTCCCGCCGCGCATCGATCGATGTAAGTGTGGCATCCATCGA 833
 70 sPheGlyHis...ThrAlaIleLeuAlaProAsnArgTrpGlnGluIlel 86
 ||||||| |||||||
 832 TTTCGGCCGCACTTCGAAGCCATCCCTTCGCCCC.....ATCA 798
 86 eusIleuIleIleSerPheIeu 93
 |||:|||||: |||
 797 CATCTATCGTAAATCGCATTTCTG 775
 seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL10148
 seq_documentation_block:
 ID ABL10148 standard; cDNA; 2969 BP.
 XX
 AC ABL10148;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 24926.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX

OS Drosophila melanogaster.
 XX
 XX WO200171042-A2.
 XX
 XX PD 27-SEP-2001.
 XX
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX
 XX PA (PEKE) PE CORP NY.
 XX
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX PI WPI; 2001-656860/75.
 DR P-PSDB; ABB66045.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS
 PS Claim 1; SEQ ID NO 24926; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 CC Sequence 2969 BP; 878 A; 613 C; 618 G; 860 T; 0 other;
 SQ
 alignment_scores:
 Quality: 70.00 Length: 41
 Ratio: 2.500 Gaps: 2
 Percent Similarity: 68.293 Percent Identity: 43.902
 alignment_block:
 US-10-048-197-2 x ABL10148/rev ..
 Align seg 1/1 to reverse of: ABL10148 from: 1 to: 2969
 54 SerArpPheGlySerAlaAspSerIleSerPheMetIleValIlely 70
 ||||||| :: |||:|||||: |||
 1882 AGTCGTTCCCGCCGCGCATCGATCGATGTAAGTGTGGCATCCATCGA 1833
 70 sPheGlyHis...ThrAlaIleLeuAlaProAsnArgTrpGlnGluIlel 86
 ||||||| |||||||
 1832 TTTCGGCCGCACTTCGAAGCCATCCCTTCGCCCC.....ATCA 1798
 86 eusIleuIleIleSerPheIeu 93
 |||:|||||: |||
 1797 CATCTATCGTAAATCGCATTTCTG 1775
 seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL12634
 seq_documentation_block:
 ID ABL12634 standard; cDNA; 3319 BP.
 XX
 AC ABL12634;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 32384.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW

[illegible]

69 elyspHeglyHsthrAlaIleu.....AlaProAsn 81
 1113 TAGGTGCAAAATAGATTTTTCATTGAAACAGCGGTAA 1162
 81 rGTpIngluIleuSerIleIleSerPheIleu 93
 1163 AATTTCAGGAAAAATTTGAACCATGTGCGCTTTTG 1200
 seq_name: /SIDSI/gc9data/geneseq/geneseqn-emb1/NA2001A.DAT:AAF28552

seq_documentation_block:
 ID AAF28552 standard; DNA: 100848 BP.

AAF28552;

04-APR-2001 (first entry)

Genomic fragment #39.

Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 bronchopulmonary; endocarditis; meningitis; ss.

Moraxella catarrhalis.

MO200078968-A2.

28-DEC-2000.

16-JUN-2000; 2000WO-US16649.

18-JUN-1999; 99US-0140121.

(INCY-) INCYTE GENOMICS INC.

Lagace RE, Patterson C, Berg KL;

WPI; 2001-041427/05.

Genomic library for identifying diagnostic and therapeutic
 compositions, and for identifying virulence factors, regulatory
 elements and drug targets, comprises Moraxella catarrhalis nucleic
 acids -

Claim 1; Page 436-459; 545pp; English.

The present invention relates to a Moraxella catarrhalis genomic library
 comprising of a combination of 41 nucleic acid molecules (see
 AAF28514-AAF28554). The library has a number of uses described in the
 specification e.g. is useful for identifying diagnostic and therapeutic
 compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 aerobic, gram-negative diplococcus, normally found among the bacterial
 flora of human upper airways. M. catarrhalis is known to cause acute,
 localised infections such as otitis media, sinusitis and bronchopulmonary
 infection and life-threatening, systemic diseases including endocarditis
 and meningitis.

Sequence 100848 BP; 28518 A; 19877 C; 22976 G; 29477 T; 0 other;

alignment_scores:

Quality: 70.00 Length: 153
 Ratio: 0.986 Gaps: 10
 Percent Similarity: 46.405 Percent Identity: 27.451

alignment_block:

US-10-048-197-2 x AAF28552 ..

Align seg 1/1 to: AAF28552 from: 1 to: 100848

13 LeuLeuThraAlaLeuValThrGlyCysValSerThr..... 25

50328 ATATGATGAGATTTTATTCGTTGATGCTGGCAACCTTATTCTACT 50377

26GlyAsnValAlaMetLysGluGlnAsnGlnInThrIleGluG 40
 50378 TGGGTGTGTATATATCTAGTCAAGATTAGACGAAATACATGCAAC 50427
 40 InThr.....IleIleLys..... 44
 50428 ATCAAGCCCAAGATTTCAATTAATAATGAATGATGCAAGCTCAACCG 50477
 45GlyLysThrAsnLysGln..... 50
 50478 TTGCCAGACAGAGCGGCAACCAACCACTAATTGTATGCGCATTCACC 50527
 51GluIleSerSerArgPheGlySer..... 58
 50528 TTTACAAATTAATAAACCACTCAGCAAGTCATTATTGGTAAAAATTGTCA 50577
 59AlaAspSerIleSerPheMetIleValValI 69
 50578 ATGCTTATGACGACGAGCAAGAGCTGGTGAATTTTATCCGTGTGGA 50627
 69 elyspHeglyHsthrAlaIleuAlaPro.....A 80
 50628 GAAA.....CATTAGATGTTACTCTCCCGTGGGAAGTTATGCTGAA 50671
 80 snArgTIPGIngluIleu.....SerLeuIleIleSerPhe 92
 50672 ATATGCTTATGAGATCTTGTGATGCGCAAGACTATTATTGGCGAAA 50721
 93 LeuTrpValLysProTyrArgProLysAsnLeuSerPheTyrLeuThrAl 109
 50722 ATACAGGTTATCCAGCGGATGAGTGTGTAATTTT...ATCACAAAT 50768
 109 aLysAla 111
 50769 CAAGGCT 50775

seq_name: /SIDSI/gc9data/geneseq/geneseqn-emb1/NA2001A.DAT:AAF55109

seq_documentation_block:

ID AAF55109 standard; DNA: 513 BP.

AAF55109;

15-MAY-2001 (first entry)

Nucleotide sequence of a BASB119 polypeptide.

BASB119; immune response; bacterial infection; genetic immunization;
 otitis media; pneumonia; sinusitis; nosocomial infection;
 invasive disease; hearing loss; fluid accumulation; antibacterial; ss.

Moraxella catarrhalis.

Key Location/Qualifiers

FT CDS 1..513

FT /tag= "a"

FT /product= "BASB119"

FT /note= "no termination codon given"

PN MO200109336-A1.

XX 08-FEB-2001.

XX 31-JUL-2000; 2000WO-EP07363.

XX 03-AUG-1999; 99GB-0018302.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Thonnard J;

XX WPI; 2001-155873/16.

DR P-PSDB; AAB67488.

seq	documentation_block:
ID	AAC85071 standard; DNA; 866 BP.
XX	
XX	
AC	AAC85071;
XX	
DT	08-MAY-2001 (first entry)
XX	
DE	Atherosclerosis-associated gene seq ID No. 7.
XX	
KW	Atherosclerosis-associated gene; stroke; myocardial infarction; human;
KW	ischemia; coronary artery disease; angina pectoris; hypertension;
KW	peripheral vascular disease; renal artery stenosis; atherosclerotic
KW	cerebroprotective; cardiact; gene therapy; hypotensive; vasotropic;
KW	antitanginal; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200104264-A2.
XX	
PD	18-JAN-2001.
XX	
PF	28-JUN-2000; 2000WO-US17887.
XX	

Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.

PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142820.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144684.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 23-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.

```

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

alignment_scores:
  Quality: 69.50      Length: 155
  Ratio: 0.952      Gaps: 8
  Percent Similarity: 47.097      Percent Identity: 23.226

```

```

alignment_block:
  US-10-048-197-2 x AAC43179 ..

```

```

Align seg 1/1 to: AAC43179 from: 1 to: 1014

```

```

3 AsnHisHisIleArgLeuThrIleSerAlaLeuLeuThr..... 15
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 AACAGAGAGATATCGGTACCGCTTCGCCGCCGCAACACTGAACCTCT 230
16 AlaLeuLeuValThrGlyCysValSer...ThrglyAsnValAla.... 29
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 CACCGTCCTCGTACCGCTCCGCGTGACAGACAGACAAATGTGTATA 280
30 .....MetlysgIugInasngInglThrIleGluInThrIleIlelys 44
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
281 AGAAATTTGAAGAGAGAGTCCGACAGTTCGTGCAAGGGGTTAGAGAG 330
45 GlyIysThrAsnlysgInglIuIleSerSerArgPheGlySerAlaAsp 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
331 ACAAAGAGAGATAGAGAGAGATCAATGAGAG.....GATGA 368
61 rIleSerPheMetIleValIleValIleIleIleIleIleIleVal 78

```

```

369 A.....:|||||:|||||:|||||:
78 lProAsnArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrp 94
|||||:|||||:|||||:|||||:|||||:
407 CTCTCT...GCTGTGAGAGGATTCATGCTTGGTCATCTTACTAGCGCT 453
95 ValIysProIyTrArgPro.....:|||||:
454 GTTCGCGAGATGAGCCCGGTTTGATCTAGTAAGAGAGAGACCTGA 503
100 .....:|||||:
504 GTTCTTTTGATGATGATGACCTTATCCAGACAGTTGACTGATGTC 553
101 .....LysAsn.....:|||||:
554 AGAAGATTCAGATAGACGCTGCTGTGATGATGATGATGATGCTTTGTT 603
107 LeuThrAlaIysAla 111
604 GTAACAGCTAAGCT 618

```

```

seq_name: /sids1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.AAT42063

```

```

seq_documentation_block:
  ID AAT42063 standard; DNA; 1830121 BP.

```

```

AAT42063;

```

```

14-SEP-1999 (first entry)

```

```

Haemophilus influenzae complete genome sequence.

```

```

Genome; Bacterium; Haemophilus influenzae; computer readable medium;
expression modulating fragment; regulation; gene expression; vector;
organism; open reading frame; ORF; ds.

```

```

Haemophilus influenzae.

```

```

W09633276-A1.

```

```

24-OCT-1996.

```

```

22-APR-1996; 96MO-US05320.

```

```

07-JUN-1995; 95US-0487429.

```

```

21-APR-1995; 95US-0426787.

```

```

07-JUN-1995; 95US-0476102.

```

```

(HUMA-) HUMAN GENOME SCI INC.

```

```

(UYJO ) UNITV JOHNS HOPKINS.

```

```

Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;

```

```

WPI; 1996-485782/48.

```

```

Haemophilus influenzae Rd genome recorded on computer readable
medium - useful for identifying commercially important nucleic acid
fragments by homology searching

```

```

Claim 1; Page 77.2-77.1091; 1291pp; English.

```

This sequence represents the complete genome sequence of the bacterium *Haemophilus influenzae* strain Rd. The invention relates to a computer readable medium (CRM) having recorded upon it the complete H. influenzae nucleotide sequence (1), a representative fragment of (1) or a nucleotide sequence at least 99% identical to (1). By providing the full-length genomic sequence in a computer readable form, it is possible to identify commercially important nucleic acid fragments and expression modulating fragments (EMFs) of the *Haemophilus* genome. The EMFs can be used to regulate the expression of a nucleic acid molecule. Vectors and altered organisms comprising the predicted ORFs can be used to produce any of the

ID AAA81456 standard; DNA; 46593 BP.
 AC AAA81456;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_4 SEQ ID NO:4.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.
 XX
 OS Neisseria meningitidis.
 PN WO200022430-A2.
 PD 20-APR-2000.
 XX
 XX
 XX 08-OCT-1999; 99WO-US23573.
 XX
 XX 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
 PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V,
 PI Rappuoli R, Pizza M;
 PL
 DR WPI: 2000-318079/27.
 XX
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisseria infections, for example, N.gonorrhoea
 XX
 PS Claim 7: Page 260-274; 1760pp; English.
 XX
 XX The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences. AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisseria bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX
 SQ Sequence 46593 BP; 11355 A; 13195 C; 11355 G; 10687 T; 1 other;

alignment_scores:
 Quality: 69.00 Length: 89
 Ratio: 1.278 Gaps: 1
 Percent Similarity: 60.674 Percent Identity: 28.090

alignment_block:
 US-10-048-197-2 x AAA81456/rev ..

Align seg 1/1 to reverse of: AAA81456 from: 1 to: 46593
 7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuValThrGlyCysAsp 23
 46127 CGCATCATCGTTCCGGCGCTTCGTTGGTGGCATTAAGCAGTTGGCG 46078
 23 IseThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
 46077 CTCATCAATTAATGTAACCGTTCCGACTAGAAACTTCAGGAACGTGCG 46028
 40 InThrIleLeuLysGlyThrAsnLysGlnGluIleSerSerArgPhe 56
 46027 CGTTGGCTTGGCGCTCAGCCCAATGCCGTAATAATCAGCAGCGCAC 45978
 57 GlySerAlaAspSerIleSerPheMetIleValValIleLysPheGlyH 73
 45977 AATGAAGGCAATCAGCATCACTTACCCCAACTGCGCTAAGCGCTGAG 45928
 73 IseThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerIle 89
 45927 CCATGCTTGTGTACCACTGTAATCAGCACAATGCGCGTTACCACTTCG 45878
 90 IleSerPheLeuTrp 94
 45877 ATGCAATTTGTTGG 45863
 seq_name: /SIDS1/gcgdata/geneseq/geneseq-n-emb1/NA2000.DAT:AAF21612
 seq_documentation_block:
 ID AAF21612 standard; DNA; 349980 BP.
 XX
 AC AAF21612;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KW ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO200066791-A1.
 PD 09-NOV-2000.
 XX
 PD 08-MAR-2000; 2000WO-US05928.
 XX
 PR 30-APR-1999; 99US-0132068.
 PR 08-OCT-1999; 99WO-US23573.
 PR 28-FEB-2000; 2000GB-0004695.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignani V,
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R,
 PI Frazer CM, Grandi G;
 XX
 DR WPI: 2000-647603/62.
 XX
 PT Neisseria meningitidis B full length genome sequence and open reading
 PT frames are used to detect, treat and prevent Neisseria infections -
 XX
 PS Claim 7: Appendix A; 692pp; English.
 XX
 CC The present invention describes the full length genome of
 CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the
 CC sequence was too long to go in a record on its own it was split into 8
 CC sequences which overlap each other at the beginning and end of each

PN WO200022430-A2

57 glySerAlaaspserIleSerPheMetIleValVal.IleLysPheGlyH 73
 ::::::::::: ||:::||| :::::::::::
 603868 AATGAAGCATACGATCACTTACCGCACTGTGGTAACGCGGTGAG 603819
 73 lSThralAlaIleuAlaProAsnArgTTPGlnIuIleuSerIeuIle 89
 ::||:::||| ::||:::|||
 603818 CCAATGCTATGTATACGATGTAATACGACATCGCGTTACCACTTCGG 603769
 90 lIleSerPheLeuTIP 94
 ::::::::::: ||:::|||
 603768 ATGCAATTGTGTTGG 603754

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:ABA03041

seq_documentation_block:
 ID ABA03041 standard; DNA: 2944528 BP.

AC ABA03041;
 XX
 DT 05-FEB-2002 (first entry)
 XX
 DE Listeria monocytogenes EGD-e genome sequence.
 XX
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease; ds.
 XX
 OS Listeria monocytogenes.
 XX
 PN WO200177335-A2.
 XX
 PD 18-OCT-2001.
 XX
 PE 11-APR-2001; 2001WO-FR01118.
 XX
 PR 11-APR-2000; 2000FR-0004629.
 XX
 PA (INSP) INSP PASTEUR.
 XX
 PI Buchrieser C, Frangeul L, Couve E, Rusanick C, Fsihl H, Dehoux P;
 PI Dussurget O, Chetoui F, Nedjati H, Glaser P, Kunst F, Cossart P;
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tlerraz-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Madueno E, De Pablo B, Weiland J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voss H;
 XX
 DR WPI; 2002-010914/01.
 XX
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides
 XX
 PS Claim 1; SEQ ID NO 1, 192pp; French.
 XX
 XS
 CC The present sequence is the genome sequence of Listeria monocytogenes
 CC EGD-e. This sequence and fragments of this sequence are useful for
 CC selecting probes and primers for detecting genes in L. monocytogenes and
 CC related organisms, and to study genetic polymorphisms and other genomes.
 CC Proteins (AB847297-AB850149) expressed from the present sequence are
 CC useful for raising specific antibodies, identification of L.
 CC monocytogenes and related organisms, and for biosynthesis and
 CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
 CC proteins encoded by it are also useful for selecting compounds that
 CC regulate gene expression and cell replication and modulate L.
 CC monocytogenes-related diseases. In addition, this sequence and proteins
 CC encoded by it are useful in pharmaceutical and vaccine compositions for
 CC the treatment or prevention of infections by L. monocytogenes and related
 CC organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX

Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;
 alignment_scores:
 Quality: 68.50 Length: 49
 Ratio: 1.851 Gaps: 1
 Percent Similarity: 75.510 Percent Identity: 34.694

alignment_block:
 US-10-048-197-2 x ABA03041/rev ..

Align seg 1/1 to reverse of: ABA03041 from: 1 to: 2944528

4 HisHisIleArgIeuThrIleSerAlaIleuLeuThrAlaIleuValTh 20
 ||||| :: ||:::||| ::||:::|||
 1863702CACCATCGCATGCGTAAATAAATCGCTTAAACGGTGAATTAATTA 1863653
 20 rGLyCysValSer...ThrGLyAsnValAlaMetLysGLuGlnAsnGln 36
 ||| ||::||| ::||:::|||
 1863652TGCATTAAGTAAATGATTCCTTTCGATGCGCTAAATAATCCAGTATCAAT 1863603
 36 lnrHrIleGluGlnThrIleIleLysGLyLysThrAsnLysGlnGlu 51
 ::||:::||| ::||:::|||
 1863602CAACGTTAAGCCATCTTCTTAATAAACAAGTAAATTCACAGGAA 1863556

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAC37886

seq_documentation_block:
 ID AAC37886 standard; DNA: 470 BP.

AC AAC37886;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 19019.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PE 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126284.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.


```
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158028.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
```

```
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161320.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

alignment_scores:
  Quality: 68.00      Length: 59
  Ratio: 1.789       Gaps: 3
  Percent Similarity: 64.407   Percent Identity: 35.593

alignment_block:
US-10-048-197-2 x AAC37886/rev ..

Align seg 1/1 to reverse of: AAC37886 from: 1 to: 470

46 LysThrAsnLysGlnGluIleSerSerArgPheGlySerAlaSpSerI 62
   ||||| ||| :|||:||||| |||:|||||:
347 AAAACCTAAAGCTTCCTGACCTCT.....TCCTGATACACT 307
62 eserPheMetIleValIleLysPhe.....GlyHisThrAlaI 76
   :|||||:|||||:|||||:|||||: ||| :||| :
306 GTCTTCATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 257
76 LeuAlaProAsnArgTyrGlnGluIleLeuSer..LeuIleIleSerP 92
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
256 TCCCGTTCATTCATGCATCACTCATTCATTCATTCATTCATTCAT 207
92 heLeuThrPValLysProTyrArgPro 100
   || ||||| ||| |||||
206 TCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 181

seq_name: /SIDSL/gcdata/geneseq/geneseqn-emb1/NA2000.DAT:AAC99183

seq_documentation_block:
ID AAC99183 standard; cDNA; 584 BP.
XX
AC AAC99183;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:411.
XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
   detection; diagnosis; identification; cytosolic; neuroprotective;
   neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
   antiinflammatory; cardiac; gene therapy; chromosome mapping;
   linkage analysis; tissue identification; tissue typing; forensic;
   neural; immune system; muscular; reproductive; gastrointestinal;
   pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS Homo sapiens.
XX
PN WO20005320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000MO-US053989.
XX
PR 12-MAR-1999; 9905-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR MPI; 2000-579444/54.
XX
P-PSDB; AAB54418.
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
```

```
PT treating, or ameliorating a medical condition, particularly pancreatic
   cancer, or for use in assays for diagnosing a pathological condition -
   Claim 1; Page 845; 1379pp; English.
XX
AC AAC98773 to AAC99231 encode the human pancreatic cancer associated
   proteins, called pancreatic cancer antigens, given in AAB54008 to
   AAB54466. The human pancreatic cancer antigens have cytosolic,
   neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
   gynaecological, cardiac and antiinflammatory activities, and can be used
   in gene therapy. The polynucleotide and proteins can be used for
   preventing, treating, or ameliorating a medical condition or in assays
   for diagnosing a pathological condition or a susceptibility to one in a
   subject. Binding partners to the proteins and the activity of the
   proteins can be identified. The pancreatic cancer antigens can be used to
   detect, treat or prevent pancreatic disorders, especially cancer.
   Agonists and antagonists to the antigens can be screened for. The
   pancreatic cancer antigen polynucleotides can be used to design nucleic
   acid hybridisation probes that can be used in chromosome mapping, linkage
   analysis, tissue identification and/or typing and a variety of forensic
   and diagnostic methods. The proteins can be used to generate antibodies
   which are used to purify, detect and target the polypeptides, including
   both in vivo and in vitro diagnostic and therapeutic methods. The
   proteins can be used to treat or prevent neural, immune system, muscular,
   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
   sequences used in the exemplification of the present invention.
XX
SQ Sequence 584 BP; 165 A; 99 C; 164 G; 148 T; 8 other;

alignment_scores:
  Quality: 68.00      Length: 84
  Ratio: 1.333       Gaps: 4
  Percent Similarity: 60.714   Percent Identity: 30.952

alignment_block:
US-10-048-197-2 x AAC99183 ..

Align seg 1/1 to: AAC99183 from: 1 to: 584

8 LeuThrIleSerAlaLeuThrAlaLeuLeuValThrGlyCysValSe 24
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
242 GTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 291
24 rThnGly.....AsnValAlaMetLysGlu.GlnAsnGlnGln 36
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
292 AGTGGGATGCTATTAATGATTTGGTAATGATGAAGCAATTTTGGAA 341
37 ThrIleGlnIleThrIleIleLysGlyLysThrAsnLysGlnGluLe 53
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
342 GTGCTGGAACCTACATGATTTGGGGAATTCACAAATCAG.....TC 385
53 rSerArgPheGlySerAlaSpSerIleSerPheMetIleValIleL 70
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
386 TTCAAAATTTTGGGACCATGAGGAGCAATTTT..... 419
70 yspheGlyHisThrAlaIleLeuAlaProAsnArgTyrGlnGluIle 86
   :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
420 .....GGAACGACAACTCTGCGCTTATGCGGCTGGAAGCAATACTT 461

seq_name: /SIDSL/gcdata/geneseq/geneseqn-emb1/NA2000.DAT:AAC35309

seq_documentation_block:
ID AAC35309 standard; DNA; 1176 BP.
XX
AC AAC35309;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 9720.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
```



```

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157573.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

alignment_scores:

```

Quality: 68.00 Length: 59
Ratio: 1.619 Gaps: 3
Percent Similarity: 71.186 Percent Identity: 28.814

```

alignment_block:

US-10-048-197-2 x AAC35309 ..

Align seq 1/1 to: AAC35309 from: 1 to: 1176

```

3 AsnHisHisIleArgLeuThrIleSerAlaLeuLeuThr..... 15
|||||:|||||: |||
263 AACGAGAGATATCGGTACCGTTCGCCGCCGCAACACTGAACCTCT 312
16 AlaLeuLeuValThrGlyCysValSer...ThrGlyAsnValAla.... 29
:|||||:|||||: |||
313 CACCGTCCTCTTACCGGTGCCGTGGAGAACGACAAATGTCGTATA 362
30 .....MetysgIugInAsnGInGInThrIleGluGInThrIleIleLys 44
:|||||:|||||: |||
363 ACMAATTTGAAGAGAGAGTGGAGCCACTGTGTCGAAGGGGTTAGTGAGG 412
45 gLyLyTThrAsnLySgInGInIleSer 53
||| :|||:|||||:
413 ACAAGAGAGATAGAGAGAAATCAAT 439

```

seq_name: /SIS1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAD21310

seq_documentation_block:
ID AAD21310 standard; cDNA; 2144 BP.

AC AAD21310;

DT 28-JUN-2002 (first entry)

DE Arabidopsis thaliana SDS cDNA.

KW Meiotic prophase I; chromosome 1; male sterile SDS mutant; apomixis;
KW plant breeding; seed production; SDS protein; ss.

OS Arabidopsis thaliana.

FH Key Location/Qualifiers
FT CDS 98..1834
FT /tag=a
FT /product="SDS protein"

PN WO200174144-A1.

XX 11-OCT-2001.

PD 29-MAR-2001; 2001WO-US09875.

PF 31-MAR-2000; 2000US-193523P.

PR (UYPE-) UNIV PENNSYLVANIA STATE.

PA Ma H;

PI WPI: 2001-662939/76.

DR P-PSDB; AAE12998.

XX A new plant gene from Arabidopsis, designated SDS, mutations in which
PT are associated with inability to produce pollen, is useful for the
PI production of male sterile plants for plant breeding

PS Claim 8; Page 39; 47pp; English.

XX The patent discloses novel plant gene from Arabidopsis thaliana,
CC designated SDS, which is associated with a failure to maintain
CC homologue attachment during meiotic prophase I. The SDS gene is
CC located on chromosome 1 and is useful in plant breeding to produce
CC male sterile SDS mutants and cloned progeny by apomixis where
CC meiosis is bypassed in seed production. The gene may also be
CC used as a probe to identify related genes in other plant species,
CC and to identify and isolate other genes of the meiosis regulatory
CC pathway. The present sequence is a cDNA encoding Arabidopsis thaliana
CC SDS protein.

SQ Sequence 2144 BP; 649 A; 461 C; 467 G; 567 T; 0 other;

alignment_scores:

```

Quality: 68.00 Length: 123
Ratio: 0.958 Gaps: 9
Percent Similarity: 57.724 Percent Identity: 31.707

```

alignment_block:

US-10-048-197-2 x AAD21310 ..

Align seq 1/1 to: AAD21310 from: 1 to: 2144

```

5 HisIleArgLeuThr.IleSerAlaLeuLeuThrAlaLeuLeuVal...T 20
|||||:|||||: |||
176 CATGTTAGAACACAGATTAACGAGTTGCCGTAATGGCTTAAGAGTGTGA 1815
20 hrGlyCysVal.serThrGlyAsnValAlaMetLySgInGIn..... 33
|||||:|||||: |||
1816 CTGTTCTCTTGGCAGTAAACCAAT.....CAAAAAGAACAAAACCCTA 1859

```

```

34  AaangInThrIle.....GlugInThrIleIleIysG 45
|||||:|||||:|||||:|||||:|||||
1860 AAACGACGACAGTATCTCCGATACCAACACAGGCTTATCTACTA 1909
45  LyIystrAsnLysGInGluIleSerSerArgPheGlySerAlaAsp... 60
|||||:|||||:|||||:|||||:|||||
1910 TTATGACAAAACAAACAGAGTAGTATAGAAAGCTCTGACAGATTAA 1959
61  .....SerIleSerPheMetI 66
|||||:|||||:|||||:|||||:|||||
1960 TATACTTAATCGAGCTGAGCTTAATAGCTCTTATACCAATTAATAG 2009
66  eValIValIleLysPheGlyHsThr.....AlaIleLeuAlaProAsn 81
|||||:|||||:|||||:|||||:|||||
2010 TGGCACCATTGTGCTGCTCATACACATTATTTCTTATTTCCCTAATT 2059
81  rGTRP...GInGluIleLeuSerLeuIleIleSerPheLeuTPValIys 96
|||||:|||||:|||||:|||||:|||||
2060 CATTAGACTCTCATATCTTAAAAAAGAAATATTTCTGTTGAAAAAAA 2109
97  ProTYArgProLys 101
::: |||
2110 AAAAAAAAAAAAAA 2124

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH17234

seq_documentation_block:
ID  AAH17234 standard; cDNA; 2303 BP.
XX
XX  AAH17234;
AC
XX  26-JUN-2001 (first entry)
DT
XX  Human cDNA sequence SEQ ID NO:16615.
DE
XX  Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX  Homo sapiens.
OS
XX  EP1074617-A2.
PN
XX  07-FEB-2001.
PD
XX  28-JUL-2000; 2000EP-0116126.
PE
XX  29-JUL-1999; 99JP-0248036.
PR  27-AUG-1999; 99JP-0300253.
PR  11-JAN-2000; 2000JP-0118776.
PR  02-MAY-2000; 2000JP-0183767.
PR  09-JUN-2000; 2000JP-0241899.
XX
XX  (HELI-) HELIX RES INST.
PA
PI  Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI  Ishii S, Sugiyama T, Makamatsu A, Nagai K, Otsuki T;
XX
XX  WPI; 2001-318749/34.
DR
XX  Primer sets for synthesizing polynucleotides, particularly the 5602
PT  full-length cDNAs defined in the specification, and for the detection
PT  and/or diagnosis of the abnormality of the proteins encoded by the
PT  full-length cDNAs -
XX
XX  Claim 8: SEQ ID 16615; 2537bp + CD ROM; English.
XX
XX  The present invention describes primer sets for synthesizing 5602
CC  full-length cDNAs defined in the specification. Where a primer set
CC  comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC  to the complementary strand of a polynucleotide which comprises one of
CC  the 5602 nucleotide sequences defined in the specification, where the
CC  oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC  of an oligonucleotide comprising a sequence complementary to the
CC  complementary strand of a polynucleotide which comprises a 5'-end

```

```

CC  sequence and an oligonucleotide comprising a sequence complementary to a
CC  polynucleotide which comprises a 3'-end sequence, where the
CC  oligonucleotide comprises at least 15 nucleotides and the combination of
CC  the 5'-end sequence/3'-end sequence is selected from those defined in
CC  the specification. The primer sets can be used in antisense therapy and
CC  in gene therapy. The primers are useful for synthesizing polynucleotides,
CC  particularly full-length cDNAs. The primers are also useful for the
CC  detection and/or diagnosis of the abnormality of the proteins encoded by
CC  the full-length cDNAs. The primers allow obtaining of the full-length
CC  cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC  AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC  AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC  represent oligonucleotides, all of which are used in the exemplification
CC  of the present invention.
XX
XX  Sequence 2303 BP; 703 A; 566 C; 502 G; 532 T; 0 other;
SQ

alignment_scores:
Quality: 68.00 Length: 124
Ratio: 1.062 Gaps: 6
Percent Similarity: 51.613 Percent Identity: 26.613

alignment_block:
US-10-048-197-2 x AAH17234/rev ..

Align seg 1/1 to reverse of: AAH17234 from: 1 to: 2303

2  LeuAsnHsHisIle.ArgLeuThrIleSerAlaLeuThrAlaLeuL 18
|||||:|||||:|||||:|||||:|||||
1151 CTATATCATCGACGTGCGAGGCTTCTGCGCTGCATACCTCTACGCGTT 1102
18  euValThrGlyCs.....ValSer 24
::: |||||
1101 CTACAAATGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1052
25  ThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGlnGln 41
|||||:|||||:|||||:|||||:|||||
1051 TTGGGGCTCTCAGAGATTGTGAGTCTGTCTCAGACTACACAGTAC 1002
41  rIleIleLysGlyStrAsnLysGInGluIleSerSerArgPheGly 58
||| ::| ::| ::| ::| ::| ::|
1001 CTGCATC...AGCTGTCTCAGCTGCGTAAGTTAGTCCCTCTTCA 955
58  eRAla..... 59
|||||:|||||:|||||:|||||:|||||
954 GTGGGATCTTTTGAATTTCTAGAGGAAGCTGAGCATCTGGAATTC 905
60  ..AspSerIleSerPheMetIleValIleLysPheGlyHsThrAl 75
|||||:|||||:|||||:|||||:|||||
904 AAGGACAGTTTTCCTGGGTGACACA.....TTTCTGA 870
75  alleLeuAlaProAsnArgTrpGInGluIleLeuSerLeuIleIleSer 92
||| |||:|||||:|||||:|||||:|||||
869 GATTTCCTTCCCAATACTGCGGTAAGATTTTATCAGAAATGTCATCT 820
92  he.LeuTrpValIysPro 97
|| ||||| |||
819 TTGCTTGGGCTCTAGCCC 802

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV52205

seq_documentation_block:
ID  AAV52205 standard; DNA; 14872 BP.
XX
XX  AAV52205;
AC
XX  23-OCT-1998 (first entry)
DT
XX  Streptococcus pneumoniae genome fragment SEQ ID NO:72.
DE
XX  Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
XX  computer readable medium; vaccine; pharmaceutical composition; ds.
KW

```

OS Streptococcus pneumoniae.
 XX
 PN WO9818931-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US19588.
 XX
 PR 31-OCT-1996; 96US-0029960.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX
 DR WPI, 1998-272225/24.
 XX
 PT Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PS pneumoniae
 XX
 PS Claim 1; Page 593-602; 1409pp: English.

CC The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment or a sequence at least 95%
 CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridise to the target sequence and
 CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC molecules whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or
 CC expression modulating fragments of the S. pneumoniae genome. Products
 CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for S. pneumoniae.
 XX
 SO Sequence 14872 BP; 4340 A; 2874 C; 3210 G; 4448 T; 0 other;

alignment_scores:
 Quality: 68.00 Length: 121
 Ratio: 1.000 Gaps: 5
 Percent Similarity: 56.198 Percent Identity: 22.314

alignment_block:
 US-10-048-197-2 x AAV52205 ..

Align seg 1/1 to: AAV52205 from: 1 to: 14872

5 HisILARGLLeuThrIleSerAlaLeuThrAla..... 16
 ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 8446 CACTGCACAACTTACACGTGGAGCCCTTATGTGCGCCAGTAGTATTGAGG 8495
 17LeuLeuValThng 21
 |||:|||||
 8496 ACATTTAGAGAGAGAGAGAAAAAATGATGCTTACTAGTATTG 8545
 21 lYcysValSerThrGlySerValAlaMetLysGlu.....GlnAsnGln 35
 ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 8546 GTTGTGGGGCGTGGCCCAAGTGTCTATTTCAAAGATTGTCAGATGAC 8595

36 GlnThrIleGluGlnThrIleIleIleGlyLysThrAsnLysGln..... 50
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 8596 GAACATTTCACGAGATTATGATTGTCAGCCGTACCAAGCAAAATGCGCA 8645
 51 GluIleSerSerArpPheGlySerAlaAspSerIleSerPheMetIleV 67
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 8646 TGACCTGAAAGCGAAGCTAGAGCGCAAAACAGTACTCAAAATT..... 8688
 67 alValIleIysPheGlyHisThrAlaIleuAlaProAsnArgTrpGln 83
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 8689GAAACGTCAGACACTGATGATCTGTCAGAGTTGAA 8721
 84 GluIleLeuSerLeuIleIleSerPheLeuTrpValLysProIlyrArgPr 100
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 8722 GAAGTGATTCCTCCGTGATGAAAGCTAC.....AAACCGAAGCTGT 8762
 100 OlyAsnLeuSer 104
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 8763 TTTCGAATGATGCT 8775

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH15139

seq_documentation_block:

ID AAH15139 standard; CDNA: 1055 BP.

AC AAH15139;

DT 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:13197.

KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

XX Claim 8; SEQ ID 13197; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in

27 NVALAIAmetlySGlucInasngtlnrhr1IEGL.....GInt 41
||||| : : : : :
1618 TGTAAACCACTCAGCGCCAAAAATGCAGTAGAAAAGTGAATAG 1667
41 hrlellelySGlyIystrAsnlys 49
: : : : :
1668 CTTACAAAAAGCGCAAGTCATCGA 1693

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.AAF18124

seq_documentation_block:
ID AAF18124 standard; DNA; 3474 BP.
XX
AC AAF18124;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 143.
XX
Human; lung cancer associated protein; neuroprotective; cytosolic;
KW cardiocative; immunomodulatory; muscular active; vulnerability;
KW gastrointestinal; nephrotoxic; antineoplastic; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
OS Homo sapiens.
XX
PN WO20055180-A2.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05918.
XX
PR 12-MAR-1999; 99US-0124270.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM.
DR WPI; 2000-587514/55.
DR P-PDB; AAB58248.
XX
PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
PS Claim 1; Page 608-609; 1425pp; English.

polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer associated proteins represented in AAB58106 - AAB58548. Lung cancer associated proteins and polynucleotide sequences, their agonists, and antagonists may have neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular active general; vulnerable; gastrointestinal general; nephrotoxic; antineoplastic; gynecological; or antibacterial activity. The invention also includes antibodies specific for the protein or polynucleotide sequences. The lung cancer associated polynucleotide sequences may be used for detection of lung cancer, chromosome identification, as chromosome markers, and for numerous other diagnostic or research purposes. The proteins may be used to treat disorders such as neural, immune, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, and proliferative disorders. The proteins may also be used in the treatment of wounds and infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB5849 are used in the course of the invention for the identification and characterisation of the polynucleotide and protein sequences.

Sequence 3474 BP; 1062 A; 692 C; 689 G; 1015 T; 16 other;

Quality: 67.50 Length: 100
Ratio: 1.298 Gaps: 4
Percent Similarity: 52.000 Percent Identity: 28.000

alignment_block:

US-10-048-197-2 x AAF18124/rev ..

Align seg 1/1 to reverse of: AAF18124 from: 1 to: 3474

```

19 ValThrcGlyCysValSerThrcGlyAsnValAlaMetLysGluGlnAsnG1 35
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
697 GTGACGAGGACGACGACGACGACGACGACGACGACGACGACGACGACG 648
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35 ngInThrlleGluGlnThrlleThrlleLysGlyLysThrsAsnLysGlnGlu 52
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
647 GCAGCGCGGCGAGGTGG.....GGAAAGGCTGAGGCGCAAAATGTC 610
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
52 leSerSerArpHeGlySerAlaAspSerIleSerPheMetIleValVal 68
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
609 TCAGTATCCATTCCACATTGTCTGTAACCTCGGCACCTGGAATTTGTC 560
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
69 IleLysPheGly_HisThrAlaIleLeuAlaPro..... 79
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
559 TCATCAGAGGAGAACATCTGAGAGGCTAGAGCCCGATGCAGGCAAGAGAG 510
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
80 ....AsnArGTrpGlnGluIleLeuSerIleIleLeuSerPheLeuTrp 94
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
509 AGAGCTGAGAGGTGG.....CTGCTGGTTAGCTCCCTCTGG 472
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 ValLysProTyrIraProLysAsnLeuSerPheTyrLeuThraLys 110
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
471 GCCCGATCTGCGCCAGAGTATGCACCAACATCCATATCTCTCGG 424
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV19980
seq_documentation_block:
ID AAV19980 standard; cDNA; 4177 BP.

```

AC AAV19980;
DT 15-JUL-1998 (first entry)
DE Full length keratinocyte growth factor-2 coding sequence.
XX
KM Keratinocyte growth factor-2; KGF-2; fibroblast growth factor-12; FGF-12;
KM keratinocyte proliferation; growth stimulator; skin strength; mucositis;
KM epidermal thickening; wound healing; inflammatory bowel disease; therapy;
KM inflammation; hair growth promoter; ds.
XX
OS Homo sapiens.
PH Key Location/Qualifiers
FT CDS 593..1219
   /*tag= a
PN MO9806844-A1.
XX
PD 19-FEB-1998.
PF 13-AUG-1997; 97WO-US14112.
XX
PR 28-FEB-1997; 97US-0039045.
PR 13-AUG-1996; 96US-0023852.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Coleman TA, Duan R, Jimenez P, Mendrick D, Moore PA;
PI NI J, Rampy MA, Ruben SM, Zhang J;
XX
DR WPI. 1998-159536/14.
DR P-PSDB; AAM53793.
XX
PT Keratinocyte growth factor-2 deletion mutants - useful to promote

```

PT or accelerate wound healing
XX
PS Disclosure: Fig 3; 251pp; English.

XX This sequence encodes the human keratinocyte growth factor-2 (KGF-2). The
CC protein is used to create the mutants of the invention, that stimulate
CC proliferation of keratinocytes. The mutants have enhanced keratinocyte
CC growth stimulating activity as compared to wild-type KGF-2 (also known as
CC fibroblast growth factor-12). They are used to stimulate growth or
CC proliferation of keratinocytes. In particular, they are used to prevent
CC or improve the appearance of wrinkles or aged skin, improving skin
CC strength, promoting epidermal thickening, reducing scarring or improving
CC healing after cosmetic surgery. The mutants are also useful for promoting
CC wound healing, especially where an individual is wound healing impaired.
CC Wounds to be treated may be surgical or excisional wounds, deep wounds
CC involving damage of the dermis and epidermis, eye tissue wounds, dental
CC tissue wounds, oral cavity wounds, diabetic, dermal, cubitus, arterial or
CC venous stasis ulcers or burns. Treatment, especially by anastomosis, of
CC wounds caused by colonic or gastrointestinal surgical procedures can also
CC be achieved through use of the KGF-2 deletion mutants. They can also be
CC used for treatment or prevention of mucositis, inflammatory bowel
CC disease, reduction of inflammation, promoting hair growth, urothelial
CC healing or tissue growth or repair in the female genital tract, or for
CC treating tissue exposed to radiation or protecting tissue to be exposed
CC to radiation (e.g. to allow an increase in radiation dosage used to treat
CC a malignancy).

XX Sequence 4177 BP; 1332 A; 816 C; 874 G; 1155 T; 0 other;

alignment_scores:
Quality: 67.50 Length: 100
Ratio: 1.298 Gaps: 4
Percent Similarity: 52.000 Percent Identity: 28.000

alignment_block:
US-10-048-197-2 x AAV19980/rev ..

Align seg 1/1 to reverse of: AAV19980 from: 1 to: 4177

```

19 ValThrcGlyCysValSerThrcGlyAsnValAlaMetLysGluGlnAsnG1 35
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
699 GTGACGAGGACGACGACGACGACGACGACGACGACGACGACGACGACG 650
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35 ngInThrlleGluGlnThrlleThrlleLysGlyLysThrsAsnLysGlnGlu 52
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
649 GCAGCGCGGCGAGGTGG.....GGAAAGGCTGAGGCGCAAAATGTC 612
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
52 leSerSerArpHeGlySerAlaAspSerIleSerPheMetIleValVal 68
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
611 TCAGTATCCATTCCACATTGTCTGTAACCTCGGCACCTGGAATTTGTC 562
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
69 IleLysPheGly_HisThrAlaIleLeuAlaPro..... 79
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
561 TCATCAGAGGAGAACATCTGAGAGGCTAGAGCCCGATGCAGGCAAGAGAG 512
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
80 ....AsnArGTrpGlnGluIleLeuSerIleIleLeuSerPheLeuTrp 94
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
511 AGAGCTGAGAGGTGG.....CTGCTGGTTAGCTCCCTCTGG 474
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 ValLysProTyrIraProLysAsnLeuSerPheTyrLeuThraLys 110
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
473 GCCCGATCTGCGCCAGAGTATGCACCAACATCCATATCTCTCGG 426
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT:AA211169

seq_documentation_block:
ID AA211169 standard; DNA; 4177 BP.

```

XX AA211169;
AC
XX
DT 03-NOV-1999 (first entry)
XX

```

Sequence 4177 BP; 1332 A; 816 C; 874 G; 1155 T; 0 other;

05-JUN-1995; 95US-0461195.

Ratio: 1.298 Gaps: 4
Percent Similarity: 52.000 Percent Identity: 28.000

alignment_block:

US-10-048-197-2 x AAF31912/rev ..

Align seg 1/1 to reverse of: AAF31912 from: 1 to: 4177

```

19 ValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnG1 35
|||||
699 GTGACGCGGACGAGACACCAAGACCAACAAACAAAGACGACGACCA 650
35 nGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnG1 52
|||||
649 GCAGCGCGGACGAGTGC.....GGAAGCGCTAGGACACAAATGTG 612
52 LeuSerArgPheGlySerAlaAspSerIleSerPheMetIleValVal 68
|||||
611 TCAGTATCCATTTCCACATTTGACTGAAACCTCTGCGACATGGAATGTC 562
69 IleLysPheGly.HisThrAlaIleLeuAlaPro.....79
|||||
561 TCATCAGAGAGACATCTGAGAGGTAAGACCCGATGCAAGCAGAGAG 512
80 ....AsnArgTyrGlnGluIleLeuSerIleIleSerPheLeuTyr 94
|||||
511 AGAGCTCAGGTGGTNG.....CTGCTGTTAGTCCCTCTGCG 474
95 ValLysProTyrArgProLysAsnLeuSerPheTyrIleuThrAlaLys 110
|||||
473 GCGGGAGTCTGCGCAGAGTGAATGCACCAATCCATTAACCTCCTCGG 426

```

seq_name: /STDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:ABL26466

seq_documentation_block:

ID ABL26466 standard; DNA; 13239 BP.

XX ABL26466;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 30871.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Claim 1; SEQ ID NO 30871; 21pp + Sequence Listing; English.

CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 13239 BP; 3611 A; 2940 C; 3072 G; 3616 T; 0 other;

alignment_scores:

Quality: 67.50 Length: 114
Ratio: 1.107 Gaps: 4
Percent Similarity: 53.509 Percent Identity: 21.930

alignment_block:

US-10-048-197-2 x ABL26466/rev ..

Align seg 1/1 to reverse of: ABL26466 from: 1 to: 13239

```

1 MetLeuAsnHisIleArgLeuThrIleSerAlaLeuThrAlaLe 17
|||||
7963 ATGCTCGGGGACATTTGTACATA.....CTACACCCACCAATT 7926
17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnA 34
|||||
7925 ACATTGCACACATGGGCCCAAGTGTGCGTAGAGTGTGCGATCTCGAT 7876
34 sngGlnThrIleGluGlnThrIleIleLys.....44
|||||
7875 CTCGTTATCTGTAGACACACTATATAGAGATCGGCAATGCCCAAGCG 7826
45 .....GlyLysThrAsnLysGlnIleSer.....53
|||||
7825 AACCCAGCGGGAGAGATCGGCGAGCGGAGTGGCGCTATCAGAGATCA 7776
54 .....SerArgPheGlySerAlaAspSerIleSerPheMetIle.... 66
|||||
7775 TCCAATTCAAGTTTACGGGTGAGGCAATTCGGTCCAATTCATGCGGT 7726
67 ....ValValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArg 81
|||||
7725 GATTTACCCACCTCCGAGAGAGATCAACCGCGTAATCGCATGATCAT 7676
82 TrpGlnGluIleLeuSerIleIleSerPheLeuTyrVal 95
|||||
7675 TGAAGTCAATCATACAAATATACGTTCTAATTCATGTTA 7634

```

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

Thu Sep 19 07:26:05 2002

OM of: US-10-048-197-2 to: GenEmbl:* out-format : pfs
 Date: Sep 18, 2002 9:39 AM

About: Results were produced by the GenCore software, version 4.5.
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framed+p2n_model -DEV=xlp
 -O/cgrr2_1/USPTO.spool/US10048197/runatc_17092002_135956_23539/app_query.fasta.1.168
 -DB=GenEmbl -GEMT=fastlap -SUFFIX=oli.rge -GAPOP=4.350
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -DELOP=6.000 -FCGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oli.rge
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USER=US10048197_@CGNL1_1-6377 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-10-048-197-2
 Query length: 111
 Database: GenEmbl:*
 Database sequences: 1797656
 Database length: 187333701
 Search time (sec): 1904.340000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
 WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_pat:AX081233	+	111.00	2095.85	2.3e-108	336	AX081233 Sequence 1 from Patent
gb_pat:AX067463	+	111.00	2056.95	3.3e-106	94750	AX067463 Sequence 38 from Pat
gb_hhg:AC099920	+	9.00	125.05	133.11	48309	AC099920 Mus musculus clone R
gb_hhg:AC103410	+	9.00	123.09	171.19	64235	AC103410 Mus musculus clone R
gb_pr:AC064841	+	9.00	122.51	184.30	69829	AC064841 Homo sapiens chromo
gb_hhg:AL670662	+	9.00	119.15	283.64	113776	AL670662 Homo sapiens chrom
gb_pr:AF241734	+	9.00	118.79	297.12	119919	AF241734 Homo sapiens PAC M1
gb_hhg:AC103115	+	9.00	116.72	387.36	161924	AC103115 Rattus norvegicus c
em_hhg_hum:AC019286	+	9.00	116.67	389.92	163134	AC019286 Homo sapiens clone
gb_pr:AC108881	+	9.00	116.67	390.12	163230	AC108881 Homo sapiens chrom
em_hhg_hum:AC058844	+	9.00	116.63	391.72	164017	AC058844 Homo sapiens chrom
gb_hhg:AC010811	+	9.00	116.57	394.72	165412	AC010811 Homo sapiens chrom
gb_hhg:AC106051	+	9.00	116.44	401.38	168577	AC106051 Homo sapiens chrom
gb_pr:AC097661	+	9.00	116.42	402.69	169197	AC097661 Homo sapiens chrom
gb_pr:AC074041	+	9.00	116.36	405.53	170548	AC074041 Mus musculus chrom
gb_hhg:AC099145	+	9.00	116.30	408.94	172176	AC099145 Rattus norvegicus c
gb_hhg:AC013798	+	9.00	116.16	416.34	175706	AC013798 Homo sapiens clone
gb_hhg:AC009290	+	9.00	116.04	422.54	178670	AC009290 Homo sapiens clone
gb_ro:AC074224	+	9.00	115.66	444.02	188989	AC074224 Mus musculus chrom
gb_hhg:AC023140	+	9.00	115.46	455.59	194575	AC023140 Homo sapiens chrom
gb_hhg:AC068980	+	9.00	115.10	477.06	204991	AC068980 Homo sapiens chrom
gb_hhg:AC093942	+	9.00	115.07	478.77	205825	AC093942 Homo sapiens chrom
gb_hhg:AC093942	+	9.00	115.07	478.77	205825	AC093942 Homo sapiens chrom
gb_hhg:AC093981	+	9.00	114.95	486.35	209518	AC093981 Mus musculus chrom
gb_pat:AX120599	+	9.00	111.75	733.23	333500	AX120599 Sequence 515 from Pat
gb_pat:AX120599	+	9.00	111.75	733.23	333500	AX120599 Sequence 515 from Pat
gb_pat:AF228279	+	8.00	141.22	15.11	293	AF228279 Glycine max clone NTN5
gb_pat:AF228279	+	8.00	140.79	17.69	312	AF228279 Glycine max clone R14
gb_in:TTU042193	+	8.00	138.96	22.37	407	TTU042193 Teosinthrips incomsequ
gb_pl:AF060192	+	8.00	137.28	27.72	519	AF060192 Glycine max putative x
gb_pl:AB027824	+	8.00	137.26	27.82	521	AB027824 Glycine max putative x
gb_pat:AX317918	+	8.00	136.73	29.74	562	AX317918 Sequence 19 from Pat
gb_pl:AY063021	+	8.00	133.91	42.73	847	AY063021 Arabidopsis thaliana R
gb_in:HAI12371	+	8.00	133.75	43.62	867	HAI12371 H. armigera mRNA for puta
gb_in:HAI12369	+	8.00	133.73	43.75	870	HAI12369 H. armigera mRNA for puta
gb_pat:AH020714	+	8.00	133.61	44.42	885	AH020714 Sequence 3 from patent
gb_str:CMS06FCD	+	8.00	132.96	48.25	972	AF396227 T7 end of clone AROA0
em_hhgo:AC043561	+	8.00	132.73	49.70	1005	AC043561 Giardia intestinalis

gb_pl:AY035150 + 8.00 132.51 51.14 1038 Arabidopsis thaliana
 em_hhgo:inv:AC046945 + 8.00 132.20 53.22 1086 AC046945 Giardia intestinalis
 gb_pl:ZF012562 + 8.00 131.92 55.12 1130 AF012562 Zamia floridiana c
 gb_pl:SCSS02A - 8.00 131.38 59.11 1223 AF156797 Mus musculus SSO2 mr
 gb_ro:AF156797 + 8.00 126.60 109.13 2449 AF156797 Mus musculus ATP s
 gb_ov:AY034614 + 8.00 126.32 113.10 2550 AY034614 Dario repto Sec61
 gb_pr:AK027492 + 8.00 126.01 117.67 2667 AK027492 Homo sapiens CDNA

seq_name: gb_pat:AX081233

seq_documentation_block:

LOCUS AX081233 336 bp DNA linear PAT 27-FEB-2001

DEFINITION Sequence 1 from Patent WO0109337.

ACCESSION AX081233

VERSION AX081233.1 GI:13170101

KEYWORDS

SOURCE Moraxella catarrhalis.

ORGANISM Moraxella catarrhalis.

REFERENCE 1 (bases 1 to 336)

AUTHORS Thonard,J.S.

TITLE Moraxella catarrhalis antigens basb122 and basb124

JOURNAL Patent: WO 0109337-A 1 08-FEB-2001;

SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)

FEATURES

source 1..336

BASE COUNT 106 a 65 c 61 g 104 t

ORIGIN

alignment_scores:

Quality: 111.00 Length: 111

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AX081233

Align seg 1/1 to: AX081233 from: 1 to: 336

1 MetLeuAmhShHsHleArgLeuThrHleSerAlaLeuThrAlaLe 17

1 ATGTTAATCATCATATTCGCTGACTATTTCCGCTTTACTGACGCAC 50

17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetGlyGluGln 34

1 TTTGGTAAACAGGTTGGTCTTCCTGCTGTAATGTTGCAATGAAGAC 100

34 snGlnGlnThrHleGluGlnThrHleGlyGlyGlyThrAsnGln 50

101 ACCAACAAACCATTTGACACACCATCATTTAAAGGCAAGACCAATTA 150

51 GlnHleSerSerThrPheGlySerAlaAspSerHleSerPheMetLe 67

151 GAGATTTCTAGTATGATTTGCTGCTGATGATCATCTTTTATGATAG 200

67 ValHleSerPheGlyHsHsHleAlaLeuAlaProAsnArgTyrGln 84

201 GGTAAATTAATTTGGACATACCGCATACCTGCTCCAAACGATGGCA 250

84 LuLeuLeuSerLeuHleHleSerPheLeuThrValLysProTyrArgPro 100

251 AAATTTTATCCCTTAATTAATTTCTTTCTTGGTGAACCATACAGACA 300

101 LysAsnLeuSerPheThrLeuThrAlaLysAla 111

301 AAGAACTTGTCAATTTTATTTTGACAGCAAGCA 333

seq_name: gb_pat:AX067463

```

seq_documentation_block:
LOCUS      AX067463          94750 bp      DNA          linear      PAT 24-JAN-2001
DEFINITION Sequence 38 from Patent WO0078968.
ACCESSION  AX067463
VERSION     AX067463.1  GI:12545083
KEYWORDS
SOURCE      Moraxella catarrhalis.
            Moraxella catarrhalis.
            Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
            Moraxella.
REFERENCE   1 (bases 1 to 94750)
            Lagace, R.E., Patterson, C. and Berg, K.L.
            Nucleotide sequences of moraxella catarrhalis genome
            Patent: WO 0078968-A 38 28-DEC-2000;
            Incyte Genomics, Inc. (US)
FEATURES
            location/Qualifiers
            source
            1..94750
            /organism="Moraxella catarrhalis"
            /db_xref="taxon:480"
BASE COUNT  27128 a 18811 c 21444 g 27367 t
ORIGIN
alignment_scores:
            Quality: 111.00      Length: 111
            Ratio: 1.000        Gaps: 0
            Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x AX067463
Align seg 1/1 to: AX067463 from: 1 to: 94750
1 MetLeuAnshHisIleAArgLeuThrIleSerAlaLeuLeuThralaIe 17
|||||
75267 ATGTTAATCATCATATTCCTGCTGACTATTCGCTTATGACGCAC 75316
|||||
17 ULeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGln 34
|||||
75317 TTGGTACAGGGTGTGTTCTCTACTGCTAATGTTGCAATGAAGACCAA 75366
|||||
34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
|||||
75367 ACCAACAAACATTGAGCAGACATCATTTAAGGCAAGACCAATTAAGCAA 75416
|||||
51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVa 67
|||||
75417 GAGATTTCTAGTAGATTGGTTCTCTGCTGATAGCATCTTTATGATAGT 75466
|||||
67 ValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTTPGInG 84
|||||
75467 GGTAAATTAATTTGGACATACCGCATCTCGCTCCAAACGATGCGCAAG 75516
|||||
84 LuIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
|||||
75517 AATTTTATCCCTTATTAATTTCTTTCTTTGGGTGAACCAATACAGACCA 75566
|||||
101 LysAsnLeuSerPheTyrLeuThralaLysAla 111
|||||
75567 AAGAACTGTCTATTTATTTGACAGCAAGCA 75599
|||||
seq_name: gb_htg:AC099920
seq_documentation_block:
LOCUS      AC099920          48309 bp      DNA          linear      HTG 22-NOV-2001
DEFINITION Mus musculus clone RP23-13M14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION  AC099920
VERSION     AC099920.1  GI:17047286
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
            Mus musculus.
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE   1 (bases 1 to 48309)
AUTHORS     Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE       Mus musculus, clone RP23-13M14
JOURNAL     Unpublished
            2 (bases 1 to 48309)
AUTHORS     Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
            Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B.,
            Brown, A., Camarata, J., Campopiano, A., Chang, J., Chararo, B.,
            Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
            Cooke, P., DeRellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
            Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
            Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N.,
            Hagos, B., Heaford, A., Horton, L., Hulme, M., Iliev, I., Johnson, R.,
            Jones, C., Kamat, A., Karatas, A., Kells, C., Labocque, K.,
            Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
            Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
            McCarthy, M., McKwan, P., McKernan, K., McPheters, R., Meldrum, J.,
            Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
            Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
            Oliver, J., Peterson, K., Phunkhang, P., Pierref, P., Pollara, V.,
            Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
            Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,
            Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
            Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
            Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
            Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
            Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
            Direct Submission
            Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            ALL repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RK/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L13528
            Center clone name: 13_M_14
            -----
            * NOTE: This record contains 61 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            1 678: contig of 678 bp in length
            679 778: gap of 100 bp
            779 1501: contig of 723 bp in length
            1502 1601: gap of 100 bp
            1602 2292: contig of 691 bp in length
            2293 2392: gap of 100 bp
            2393 3082: contig of 690 bp in length
            3083 3182: gap of 100 bp
            3183 3831: contig of 649 bp in length
            3832 3931: gap of 100 bp
            3932 4480: contig of 549 bp in length
            4481 4580: gap of 100 bp
            4581 5283: contig of 703 bp in length
            5284 5383: gap of 100 bp
            5384 6063: contig of 680 bp in length
            6064 6163: gap of 100 bp
            6164 6896: contig of 733 bp in length
            6897 6996: gap of 100 bp
            6997 7641: contig of 645 bp in length
            7642 7741: gap of 100 bp

```



```

7742      8458: contig of 717 bp in length
8459      8558: gap of 100 bp
8559      9290: contig of 732 bp in length
9291      9390: gap of 100 bp
9391      10112: contig of 722 bp in length
10113      10212: gap of 100 bp
10213      10927: contig of 715 bp in length
10928      11027: gap of 100 bp
11028      11736: contig of 709 bp in length
11737      11836: gap of 100 bp
11837      12561: contig of 725 bp in length
12562      12661: gap of 100 bp
12662      13345: contig of 684 bp in length
13346      13445: gap of 100 bp
13446      14128: contig of 683 bp in length
14129      14228: gap of 100 bp
14229      14901: contig of 673 bp in length
14902      15001: gap of 100 bp
15002      15681: contig of 680 bp in length
15682      15781: gap of 100 bp
15782      16458: contig of 677 bp in length
16459      16558: gap of 100 bp
16559      17233: contig of 675 bp in length
17234      17333: gap of 100 bp
17334      18010: contig of 677 bp in length
18011      18110: gap of 100 bp
18111      18832: contig of 722 bp in length
18833      18932: gap of 100 bp
18933      19663: contig of 731 bp in length
19664      19763: gap of 100 bp
19764      20436: contig of 673 bp in length
20437      20536: gap of 100 bp
20537      21242: contig of 706 bp in length
21243      21342: gap of 100 bp
21343      22045: contig of 703 bp in length
22046      22145: gap of 100 bp
22146      22842: contig of 697 bp in length
22843      22942: gap of 100 bp
22943      23652: contig of 710 bp in length
23653      23752: gap of 100 bp
23753      24481: contig of 729 bp in length
24482      24581: gap of 100 bp
24582      25312: contig of 731 bp in length
25313      25412: gap of 100 bp
25413      26099: contig of 687 bp in length
26100      26199: gap of 100 bp
26200      26890: contig of 691 bp in length
26891      26990: gap of 100 bp
26991      27663: contig of 673 bp in length
27664      27763: gap of 100 bp
27764      28454: contig of 691 bp in length
28455      28554: gap of 100 bp
28555      29267: contig of 713 bp in length
29268      29367: gap of 100 bp
29368      30081: contig of 714 bp in length
30082      30181: gap of 100 bp
30182      30867: contig of 686 bp in length
30868      30967: gap of 100 bp
30968      31697: contig of 730 bp in length
31698      31797: gap of 100 bp
31798      32483: contig of 686 bp in length
32484      32583: gap of 100 bp
32584      33245: contig of 662 bp in length
33246      33345: gap of 100 bp
33346      34026: contig of 681 bp in length
34027      34126: gap of 100 bp
34127      34845: contig of 719 bp in length
34846      34945: gap of 100 bp
34946      35647: contig of 702 bp in length
35648      35747: gap of 100 bp
35748      36423: contig of 676 bp in length
36424      36523: gap of 100 bp
36524      37241: contig of 718 bp in length

```

```

* 37242 37341: gap of 100 bp
* 37342 38019: contig of 678 bp in length
* 38020 38119: gap of 100 bp
* 38120 38780: contig of 661 bp in length
* 38781 38880: gap of 100 bp
* 38881 39596: contig of 716 bp in length
* 39597 39696: gap of 100 bp
* 39697 40393: contig of 697 bp in length
* 40394 40493: gap of 100 bp
* 40494 41169: contig of 676 bp in length
* 41170 41269: gap of 100 bp
* 41270 41922: contig of 653 bp in length
* 41923 42022: gap of 100 bp
* 42023 42711: contig of 689 bp in length
* 42712 42811: gap of 100 bp
* 42812 43507: contig of 696 bp in length
* 43508 43607: gap of 100 bp
* 43608 44336: contig of 729 bp in length
* 44337 44436: gap of 100 bp
* 44437 45109: contig of 673 bp in length
* 45110 45209: gap of 100 bp
* 45210 45906: contig of 697 bp in length
* 45907 46006: gap of 100 bp
* 46007 46723: contig of 717 bp in length
* 46724 46823: gap of 100 bp
* 46824 47513: contig of 690 bp in length
* 47514 47613: gap of 100 bp
* 47614 48309: contig of 696 bp in length.
  Location/Qualifiers
    source          1..48309
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /clone="RP23-13M14"
                    /clone_1lb="RPI-23 Female Mouse BAC"
BASE COUNT      13627 a 7821 c 7946 g 12849 t 6066 others
ORIGIN
FEATURES
  source
    alignment_scores:
      Quality: 9.00      Length: 9
      Ratio: 1.000      Gaps: 0
      Percent Similarity: 100.000      Percent Identity: 100.000
  alignment_block:
    US-10-048-197-2 x AC099920/rev ..
  Align seg 1/1 to reverse of: AC099920 from: 1 to: 48309
    84 GtutleuSerleuIleSerPhe 92
    |||||||||||||||||||
    40205 GAGATCCTGTTTGATATTTCTTT 40179
seq_name: gb_hcg:AC103410
seq_documentation_block:
LOCUS      AC103410      64235 bp      DNA      linear      HTG 24-NOV-2001
DEFINITION Mus musculus clone RP24-363H11, LOW-PASS SEQUENCE SAMPLING.
ACCESSION  AC103410
VERSION    AC103410.1 GI:17063285
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 64235)
            Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Mus musculus, clone RP24-363H11
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 64235)
            Nusbaum,C., Lander,E., All,A., Allen,N.,
            Anderson,S., Barne,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campoliano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,

```

TITLE
JOURNAL
COMMENT

Cooke, P., Dearlano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Gilde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Lacombe, K.,
Lamarez, R., Landers, T., Lejoczky, J., Levine, R., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunhamp, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Resfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (24-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L19943

Center clone name: 363_R11

* NOTE: This record contains 81 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 676 775: contig of 675 bp in length
* 776 1459: contig of 684 bp in length
* 1460 1559: gap of 100 bp
* 1560 2263: contig of 704 bp in length
* 2264 2363: gap of 100 bp
* 2364 3159: contig of 636 bp in length
* 3060 3159: gap of 100 bp
* 3160 3796: contig of 637 bp in length
* 3797 3896: gap of 100 bp
* 3897 4596: contig of 700 bp in length
* 4597 4696: gap of 100 bp
* 4697 5396: contig of 700 bp in length
* 5397 5496: gap of 100 bp
* 5497 6193: contig of 697 bp in length
* 6194 6293: gap of 100 bp
* 6294 6981: contig of 688 bp in length
* 6982 7081: gap of 100 bp
* 7082 7784: contig of 703 bp in length
* 7785 7884: gap of 100 bp
* 7885 8593: contig of 715 bp in length
* 8600 8699: gap of 100 bp
* 8700 9393: contig of 694 bp in length
* 9394 9493: gap of 100 bp
* 9494 10205: contig of 712 bp in length
* 10206 10305: gap of 100 bp
* 10306 11012: contig of 707 bp in length
* 11013 11112: gap of 100 bp
* 11113 11807: contig of 695 bp in length

* 11808 11907: gap of 100 bp
* 11908 12604: contig of 697 bp in length
* 12605 12704: gap of 100 bp
* 12705 13390: contig of 686 bp in length
* 13391 13490: gap of 100 bp
* 13491 14158: contig of 668 bp in length
* 14159 14258: gap of 100 bp
* 14259 14953: contig of 695 bp in length
* 14954 15053: gap of 100 bp
* 15054 15749: contig of 696 bp in length
* 15750 15849: gap of 100 bp
* 15850 16530: contig of 681 bp in length
* 16531 16630: gap of 100 bp
* 16631 17333: contig of 703 bp in length
* 17334 17433: gap of 100 bp
* 17434 18135: contig of 702 bp in length
* 18136 18235: gap of 100 bp
* 18236 18908: contig of 673 bp in length
* 18909 19008: gap of 100 bp
* 19009 19717: contig of 709 bp in length
* 19718 19817: gap of 100 bp
* 19818 20503: contig of 686 bp in length
* 20504 20603: gap of 100 bp
* 20604 21290: contig of 687 bp in length
* 21291 21390: gap of 100 bp
* 21391 22078: contig of 688 bp in length
* 22079 22178: gap of 100 bp
* 22179 22888: contig of 710 bp in length
* 22889 22988: gap of 100 bp
* 22989 23704: contig of 716 bp in length
* 23705 23804: gap of 100 bp
* 23805 24513: contig of 709 bp in length
* 24514 24613: gap of 100 bp
* 24614 25308: contig of 695 bp in length
* 25309 25408: gap of 100 bp
* 25409 26117: contig of 709 bp in length
* 26118 26217: gap of 100 bp
* 26218 26908: contig of 691 bp in length
* 26909 27008: gap of 100 bp
* 27009 27695: contig of 687 bp in length
* 27696 27795: gap of 100 bp
* 27796 28490: contig of 695 bp in length
* 28491 28590: gap of 100 bp
* 28591 29286: contig of 666 bp in length
* 29287 29386: gap of 100 bp
* 29387 30092: contig of 706 bp in length
* 30093 30192: gap of 100 bp
* 30193 30896: contig of 704 bp in length
* 30897 30996: gap of 100 bp
* 30997 31697: contig of 701 bp in length
* 31698 31797: gap of 100 bp
* 31798 32513: contig of 716 bp in length
* 32514 32613: gap of 100 bp
* 32614 33283: contig of 670 bp in length
* 33284 33383: gap of 100 bp
* 33384 34068: contig of 685 bp in length
* 34069 34168: gap of 100 bp
* 34169 34809: contig of 641 bp in length
* 34810 34909: gap of 100 bp
* 34910 35591: contig of 682 bp in length
* 35592 35691: gap of 100 bp
* 35692 36381: contig of 690 bp in length
* 36382 36481: gap of 100 bp
* 36482 37178: contig of 697 bp in length
* 37179 37278: gap of 100 bp
* 37279 37973: contig of 695 bp in length
* 37974 38073: gap of 100 bp
* 38074 38759: contig of 686 bp in length
* 38760 38859: gap of 100 bp
* 38860 39558: contig of 699 bp in length
* 39559 39658: gap of 100 bp
* 39659 40336: contig of 678 bp in length
* 40337 40436: gap of 100 bp

```

* 40437 41136: contig of 700 bp 1n length
* 41137 41236: gap of 100 bp
* 41237 41882: contig of 646 bp 1n length
* 41883 41982: gap of 100 bp
* 41983 42683: contig of 701 bp 1n length
* 42684 42783: gap of 100 bp
* 42784 43485: contig of 702 bp 1n length
* 43486 43585: gap of 100 bp
* 43586 44275: contig of 690 bp 1n length
* 44276 44375: gap of 100 bp
* 44376 45066: contig of 691 bp 1n length
* 45067 45166: gap of 100 bp
* 45167 45847: contig of 681 bp 1n length
* 45848 45947: gap of 100 bp
* 45948 46652: contig of 705 bp 1n length
* 46653 46752: gap of 100 bp
* 46753 47468: contig of 716 bp 1n length
* 47469 47568: gap of 100 bp
* 47569 48268: contig of 700 bp 1n length
* 48269 48368: gap of 100 bp
* 48369 48066: contig of 698 bp 1n length
* 49067 49166: gap of 100 bp
* 49167 49845: contig of 679 bp 1n length
* 49846 49945: gap of 100 bp
* 49946 50630: contig of 685 bp 1n length
* 50631 50730: gap of 100 bp
* 50731 51442: contig of 712 bp 1n length
* 51443 51542: gap of 100 bp
* 51543 52232: contig of 690 bp 1n length
* 52233 52332: gap of 100 bp
* 52333 53031: contig of 699 bp 1n length
* 53032 53131: gap of 100 bp
* 53132 53826: contig of 695 bp 1n length
* 53827 53926: gap of 100 bp
* 53927 54625: contig of 699 bp 1n length
* 54626 54725: gap of 100 bp

```

alignment_scores:

```

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

```

US-10-048-197-2 x AC103410 ..

```

```

Align seg 1/1 to: AC103410 from: 1 to: 64235

```

```

      8 leuThrIleSerAlaLeuThrAla 16
      |||||
      49572 TTGACCATTTGCGCGTTGTTAACGCGT 49598

```

```

seq_name: gb_pr:AC064841

```

```

seq_documentation_block:

```

```

LOCUS AC064841 69829 bp DNA linear PRI 01-FEB-2002

```

```

DEFINITION Homo sapiens chromosome 2 clone CTD-2015A10, complete sequence.

```

```

ACCESSION AC064841.3 GI:18464243

```

```

VERSION AC064841.3 GI:18464243

```

```

KEYWORDS HTG.

```

```

SOURCE human.

```

```

ORGANISM Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE 1 (bases 1 to 69829)

```

```

AUTHORS Waterston,R.H.

```

```

TITLE The sequence of Homo sapiens clone

```

```

JOURNAL Unpublished

```

```

AUTHORS 2 (bases 1 to 69829)

```

```

TITLE Direct Submission

```

```

JOURNAL Submitted (23-APR-2000) Genome Sequencing Center, Washington

```

```

MO 63108, USA

```

```

REFERENCE 3 (bases 1 to 69829)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```

COMMENT On Feb 1, 2002 this sequence version replaced g1:18151025.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
Project information
Center project name: H_MS2015A10
-----
FEATURES
source
1. .69829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="CTD-2015A10"
BASE COUNT 20182 a 12890 c 12669 g 24088 t
ORIGIN

```

```

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-10-048-197-2 x AC064841 ..

```

```

Align seg 1/1 to: AC064841 from: 1 to: 69829

```

```

      86 leuSerIleIleSerPheLeuTrp 94
      |||||
      5553 CTTTCATTATATATCATTTCTTTGG 5579

```

```

seq_name: gb_hcg:AL670662

```

```

seq_documentation_block:

```

```

LOCUS AL670662 113776 bp DNA linear HTG 01-FEB-2002

```

```

DEFINITION Homo sapiens chromosome 6 clone XHbac-93B10, *** SEQUENCING IN

```

```

PROGRESSES ***, 16 unordered pieces.

```

```

ACCESSION AL670662

```

```

VERSION AL670662.3 GI:18491449

```

```

KEYWORDS HTG; HTGS_PHASE1.

```

```

SOURCE human.

```

```

ORGANISM Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE 1 (siles)

```

```

AUTHORS Sims,S.

```

```

TITLE Direct Submission

```

```

JOURNAL Submitted (31-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,

```

```

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

```

```

humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

```

```

On Feb 4, 2002 this sequence version replaced g1:18477451.

```

```

----- Genome Center -----

```

```

Center: Wellcome Trust Sanger Institute

```

```

Center code: SC

```

```

Web site: http://www.sanger.ac.uk

```

```

Contact: humquerry@sanger.ac.uk

```

```

----- Project Information

```

```

Center project name: bpg93B10

```

```

----- Summary Statistics

```

```

Assembly program: XGAP4; Version 4.5

```

```

Sequencing vector: plasmid; L08752; 100% of reads

```

```

Chemistry: Dye-terminator Big Dye; 100% of reads

```

```

Consensus quality: 107323 bases at least Q40

```

Consensus quality: 110694 bases at least Q20
 Insert size: 112276; sum-of-contigs
 Insert size: 118454; 4.0% error; agarose-1p
 Quality coverage: 4.94x in Q20 bases; sum-of-contigs quality
 coverage: 4.73x in Q20 bases; agarose-1p

NOTE: This is a 'working draft' sequence. It currently
 consists of 16 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

```

1      2288: contig of 2288 bp in length
*      2289 2388: gap of 100 bp
*      5334 5433: contig of 2945 bp in length
*      5434 5433: gap of 100 bp
*      11934 12033: contig of 6500 bp in length
*      12034 20810: contig of 8777 bp in length
*      20811 20910: gap of 100 bp
*      20911 29466: contig of 8556 bp in length
*      29467 29566: gap of 100 bp
*      32567 32572: contig of 3006 bp in length
*      32573 32672: gap of 100 bp
*      36441 36540: contig of 3768 bp in length
*      36541 38877: contig of 2337 bp in length
*      38878 38977: gap of 100 bp
*      38978 43509: contig of 4532 bp in length
*      43510 43609: gap of 100 bp
*      43610 57681: contig of 14072 bp in length
*      57682 57781: gap of 100 bp
*      57782 72280: contig of 14499 bp in length
*      72281 72380: gap of 100 bp
*      72381 86911: contig of 14531 bp in length
*      86912 87011: gap of 100 bp
*      87012 94994: contig of 7983 bp in length
*      94995 95094: gap of 100 bp
*      95095 96312: contig of 3218 bp in length
*      96313 98412: gap of 100 bp
*      98413 106958: contig of 8546 bp in length
*      106959 107059: gap of 100 bp
*      107059 113776: contig of 6718 bp in length.
  
```

FEATURES

Location/Qualifiers

```

1. .113776
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="6"
  /clone="XXbac-93B10"
  /clone_1bp="CHOR1-501"
  1. .2288
    /note="assembly_fragment:00049"
    fragment_chain:1
      clone_end:T7
      vector_side:left
      2389. .5333
        /note="assembly_fragment:00859"
        fragment_chain:1
          5434. .11933
            /note="assembly_fragment:01130"
            fragment_chain:1
              12034. .20810
                /note="assembly_fragment:00562"
                fragment_chain:2
                  20911. .29466
                    /note="assembly_fragment:00633"
                    fragment_chain:2
                      29567. .32572
                        /note="assembly_fragment:00674"
                        fragment_chain:2
                          32673. .36440
  
```

```

misc_feature /note="assembly_fragment:00655"
              fragment_chain:2
              36541. .38877
                /note="assembly_fragment:00599"
                fragment_chain:3
                38978. .43509
                  /note="assembly_fragment:00313"
                  fragment_chain:3
                  43610. .57681
                    /note="assembly_fragment:00838"
                    fragment_chain:3
                    57782. .72280
                      /note="assembly_fragment:00179"
                      fragment_chain:4
                      72381. .86911
                        /note="assembly_fragment:00755"
                        fragment_chain:4
                        87012. .94994
                          /note="assembly_fragment:00446"
                          fragment_chain:5
                          95095. .98312
                            /note="assembly_fragment:01230"
                            fragment_chain:5
                            98413. .106958
                              /note="assembly_fragment:00654"
                              107059. .113776
                                /note="assembly_fragment:01253"
                                clone_end:SP6
                                vector_side:right
  
```

BASE COUNT 30735 a 25968 c 25314 g 30248 t 1511 others

alignment_scores: Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AL670662 ..

Align seg 1/1 to: AL670662 from: 1 to: 113776

```

101 LysAsnLeuSerPheTyrLeuThrAla 109
|||||
64868 AAAAAGCTCAGTTCTACTTGACAGCA 64894
  
```

seq_name: gb-pr:AF241734

seq_documentation_block:

```

LOCUS AF241734 119919 bp DNA linear PRI 05-APR-2000
DEFINITION Homo sapiens PAC M1034 chromosome X map Xp11.4, complete sequence.
ACCESSION AF241734
VERSION AF241734.1 GI:7417369
KEYWORDS HTG.
SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 119919)
    Blechschmidt,K., Nyakatura,G., Menzel,U., Baumgart,C., Dette,M.,
    Jahn,N., Strom,T.M., Hellebrand,H., Meindl,A. and Rosenthal,A.
    Direct Submission
    Submitted (03-MAR-2000) Genome Analysis, Institute of Molecular
    Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
  
```

TITLE JOURNAL
 Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany

FEATURES

```

1. .119919
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="X"
  /map="Xp11.4"
  /clone="PAC M1034"
  repeat_region 1221. .1351
  
```

```
/note="(CAT)n"
/rpt_family="Simple_repeat"
exon 1679..1785
/note="GRAIL"
/evidence-not_experimental
complement(3238..3486)
exon /note="GRAIL"
/evidence-not_experimental
complement(3247..3329)
exon /note="XPOUND"
/evidence-not_experimental
3715..3857
/note="WZER"
/evidence-not_experimental
complement(4848..5069)
exon /note="GRAIL"
/evidence-not_experimental
complement(4890..5039)
exon /note="WZER"
/evidence-not_experimental
complement(4890..5069)
exon /note="GENSCAN"
complement(5114..5216)
exon /note="XPOUND"
/evidence-not_experimental
complement(5119..5249)
repeat_region /note="(GA)n"
/rpt_family="Simple_repeat"
complement(5438..5615)
exon /rpt_family="LTR16A"
complement(5655..5735)
/note="GRAIL"
/evidence-not_experimental
5781..5898
/rpt_family="MLT11"
complement(5941..6025)
repeat_region /rpt_family="MERS8A"
complement(6370..6471)
exon /note="GRAIL"
/evidence-not_experimental
complement(6370..6471)
exon /note="GENSCAN"
6373..6474
/note="GRAIL"
/evidence-not_experimental
complement(6589..6652)
exon /note="XPOUND"
/evidence-not_experimental
6911..7196
repeat_region /rpt_family="LTR16A"
7223..7995
/rpt_family="L1M4orf2"
7847..8729
repeat_region /rpt_family="L1M8"
8730..8799
/rpt_family="LTR16A"
complement(9152..9290)
exon /note="GRAIL"
/evidence-not_experimental
complement(9577..9684)
repeat_region /rpt_family="L2b"
9901..10097
/rpt_family="MERS3"
10692..10866
repeat_region /rpt_family="L2a"
10756..10866
repeat_region /rpt_family="L2b"
complement(11684..11746)
exon /note="GRAIL"
/evidence-not_experimental
complement(12928..13387)
repeat_region /rpt_family="MERS5A"
12973..13321
/rpt_family="AlusX"
13276..13386
/note="WZER"
/evidence-not_experimental
13568..13666
repeat_region /note="(GGA)n"
/rpt_family="Simple_repeat"
14381..14523
/rpt_family="MERS5A"
14533..14565
repeat_region /rpt_family="Charlie4"
15505..15817
repeat_region /rpt_family="AlusX"
complement(16868..16963)
repeat_region /rpt_family="MLT11"
complement(17112..17593)
exon /rpt_family="L2"
17967..18107
/note="GRAIL"
/evidence-not_experimental
complement(17967..18099)
exon /note="GRAIL"
/evidence-not_experimental
18005..18291
repeat_region /rpt_family="L1M86"
18005..18256
repeat_region /rpt_family="L1M82"
18163..18253
exon /note="GRAIL"
/evidence-not_experimental
18309..18559
repeat_region /rpt_family="L1M83"
18806..18897
/note="GRAIL"
/evidence-not_experimental
complement(18918..19221)
exon /rpt_family="AlusX"
20125..20201
repeat_region /rpt_family="LTR33"
20444..22837
repeat_region /rpt_family="LTR33"
complement(20514..20871)
repeat_region /rpt_family="THE1B"
complement(20872..22434)
exon /rpt_family="THE1-INTERNAL"
21254..21350
/note="GENSCAN"
complement(22026..22156)
exon /note="WZER"
/evidence-not_experimental
complement(22041..22247)
exon /note="GRAIL"
/evidence-not_experimental
complement(22435..22792)
repeat_region /rpt_family="THE1B"
complement(23324..23424)
exon /rpt_family="MIR"
23464..23516
/note="GRAIL"
/evidence-not_experimental
23464..23522
exon /note="GENSCAN"
complement(23649..23713)
repeat_region /rpt_family="MSTC"
complement(23735..25349)
repeat_region /rpt_family="MST-INTERNAL"
24441..24528
exon /note="GRAIL"
/evidence-not_experimental
complement(25350..25742)
repeat_region /rpt_family="MLT1A1"
25430..25529
/note="WZER"
```

```

/evidence=not_experimental
repeat_region complement(26059, 27233)
/rpt_family="L1M6"
repeat_region complement(26414, 26707)
/rpt_family="Aluub"
repeat_region 27233, 27595
/rpt_family="L1TR16C"
repeat_region 28089, 28802
/rpt_family="L1M3"
repeat_region complement(28911, 29028)
/rpt_family="Aluuo"
exon 28914, 29057
/note="M2EF"
/evidence=not_experimental
repeat_region complement(29133, 29404)
/rpt_family="L1TR16A"
repeat_region 29635, 29796
/rpt_family="M1R"
repeat_region complement(29991, 30286)
/rpt_family="AluSx"
repeat_region complement(30538, 30579)
/note="(CAAAA)n"
exon /rpt_family="Simple-repeat"
complement(30917, 31027)
/evidence=not_experimental
repeat_region 31215, 31363
/rpt_family="MERSA"
exon complement(31310, 31450)
/note="GRAIL"
/evidence=not_experimental

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AF241734 ..
Align seg 1/1 to: AF241734 from: 1 to: 119919
101 LysAsnLeuSerPheTyrLeuThrAla 109
|||||
39908 AAAAACCCTGACTTCTGACACGA 39934
seq_name: gb_hgtc:AC103115

seq_documentation_block:
LOCUS AC103115 161924 bp DNA linear HTG 21-DEC-2001
DEFINITION Rattus norvegicus clone CH230-188D19, *** SEQUENCING IN PROGRESS
***, 64 unordered pieces.
ACCESSION AC103115
VERSION AC103115.2 GI:17974602
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 161924)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
Aishbrook,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbard,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dedereich,D.A., Delaney,K.R., Delgado,O.,
Dean,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Roches,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Ieal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lien,C., Liu,J., Liu,W.,
Lousaged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,J., Liu,W.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokemto,S.,
Ogulu,M., Okunou,G., Oragune,N., Oviado,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pui,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Rutiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitani,N.,
Slisdon,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Vinson,R.,
Washington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 161924)
Morley,K.C.
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 21, 2001 this sequence version replaced gi:17062758.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJF7
Center clone name: CH230-188D19
----- Summary Statistics
Assembly Program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 132307 bases at least Q40
Consensus quality: 140655 bases at least Q30
Consensus quality: 147116 bases at least Q20
Estimated insert size: 131552; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 64 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 7625: contig of 7625 bp in length
* 7626
* 7726: gap of unknown length
* 7726
* 14756: contig of 7030 bp in length
* 14756
* 14855: gap of unknown length
* 14856
* 23333: contig of 8478 bp in length
* 23333
* 23433: gap of unknown length
* 23434
* 27343: contig of 3910 bp in length
* 27344
* 27443: gap of unknown length
* 27444
* 30836: contig of 3393 bp in length

```

30837 30936: gap of unknown length
* 30937 35060: contig of 4124 bp in length
* 35061 35160: gap of unknown length
* 35161 40254: contig of 5094 bp in length
* 40255 40354: gap of unknown length
* 40355 44541: contig of 4187 bp in length
* 44542 44641: gap of unknown length
* 44642 49051: contig of 4410 bp in length
* 49052 49151: gap of unknown length
* 49152 52529: contig of 3378 bp in length
* 52530 52629: gap of unknown length
* 52630 56269: contig of 3640 bp in length
* 56270 56369: gap of unknown length
* 56370 59272: contig of 2903 bp in length
* 59273 59372: gap of unknown length
* 59373 63381: contig of 4009 bp in length
* 63382 63481: gap of unknown length
* 63482 66550: contig of 3069 bp in length
* 66551 66650: gap of unknown length
* 66651 66857: contig of 2207 bp in length
* 66858 68958: gap of unknown length
* 68959 73301: contig of 4343 bp in length
* 73301 73400: gap of unknown length
* 73401 75764: contig of 2364 bp in length
* 75765 75864: gap of unknown length
* 75865 78928: contig of 3064 bp in length
* 78929 79028: gap of unknown length
* 79029 81347: contig of 2319 bp in length
* 81348 81448: gap of unknown length
* 81449 83030: contig of 1583 bp in length
* 83031 83131: gap of unknown length
* 83132 85382: contig of 2252 bp in length
* 85383 85482: gap of unknown length
* 85483 87766: contig of 2284 bp in length
* 87767 87866: gap of unknown length
* 87867 89888: contig of 2022 bp in length
* 89889 89988: gap of unknown length
* 89989 91867: contig of 1878 bp in length
* 91867 91967: gap of unknown length
* 91967 94685: contig of 2719 bp in length
* 94686 94785: gap of unknown length
* 94786 96401: contig of 1616 bp in length
* 96402 96501: gap of unknown length
* 96502 98769: contig of 2268 bp in length
* 98770 98869: gap of unknown length
* 98870 100269: contig of 1400 bp in length
* 100270 100369: gap of unknown length
* 100370 101921: contig of 1552 bp in length
* 101922 102021: gap of unknown length
* 102022 103110: contig of 1089 bp in length
* 103111 103210: gap of unknown length
* 103211 104666: contig of 1456 bp in length
* 104667 104766: gap of unknown length
* 104767 107055: contig of 2289 bp in length
* 107056 107155: gap of unknown length
* 107156 108844: contig of 1789 bp in length
* 108845 109044: gap of unknown length
* 109045 109045: gap of unknown length
* 109046 110910: contig of 1866 bp in length
* 110911 112762: contig of 1752 bp in length
* 112763 112862: gap of unknown length
* 112863 114085: contig of 1223 bp in length
* 114086 114185: gap of unknown length
* 114186 115740: contig of 1535 bp in length
* 115741 115840: gap of unknown length
* 115841 117251: contig of 1411 bp in length
* 117252 117351: gap of unknown length
* 117352 119708: contig of 2357 bp in length
* 119709 119808: gap of unknown length
* 119809 121548: contig of 1740 bp in length
* 121549 121648: gap of unknown length
* 121649 122930: contig of 1282 bp in length
* 122931 123030: gap of unknown length

123031 125145: contig of 2115 bp in length
* 125146 125245: gap of unknown length
* 125246 127372: contig of 2127 bp in length
* 127373 127472: gap of unknown length
* 127473 129619: contig of 2147 bp in length
* 129620 129719: gap of unknown length
* 129720 131383: contig of 1664 bp in length
* 131384 131484: gap of unknown length
* 131485 132714: contig of 1231 bp in length
* 132715 132814: gap of unknown length
* 132815 134391: contig of 1577 bp in length
* 134392 134491: gap of unknown length
* 134492 135742: contig of 1251 bp in length
* 135743 135842: gap of unknown length
* 135843 137997: contig of 2155 bp in length
* 137998 138097: gap of unknown length
* 138098 140078: contig of 1981 bp in length
* 140079 140178: gap of unknown length
* 140179 141589: contig of 1411 bp in length
* 141590 141690: gap of unknown length
* 141691 143025: contig of 1336 bp in length
* 143026 143125: gap of unknown length
* 143126 144267: contig of 1142 bp in length
* 144268 144367: gap of unknown length
* 144368 145839: contig of 1472 bp in length
* 145840 145939: gap of unknown length
* 145940 147839: contig of 1900 bp in length
* 147840 147939: gap of unknown length
* 147940 149476: contig of 1537 bp in length
* 149477 149576: gap of unknown length

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AC103115/rev ..

Align seg 1/1 to reverse of: AC103115 from: 1 to: 161924

85 ILeuSerLeuIleIleSerPheIeu 93
|||||
20164 ATTTTAAGCTTATTATATCGTTCTT 20138

seq_name: em_hcg_hum:AC019286

seq_documentation_block:
ID AC019286 standard; DNA; HTG; 163134 BP.

AC AC019286;

SV AC019286.4

DT 02-JAN-2000 (Rel. 62, Created)

DT 05-APR-2000 (Rel. 63, last updated, Version 4)

DE Homo sapiens clone RP11-21D3, WORKING DRAFT SEQUENCE, 16 unordered pieces.

HTG; HTGS_DRAFT; HTGS_PHASE1.

OS Homo sapiens (human)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia;

OC Eutheria; Primates; Catarrhini; Homidae; Homo.

RA Birren B., Linton L., Nusbaum C., Lander E.;

RT "Homo sapiens, clone RP11-21D3";

RP 1-163134

RA Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
 RA Anderson S., Baldwin J., Barna N., Beckerly R., Beda F., Boguslavsky L.,
 RA Boughalter B., Brown A., Burkett G., Castle A., Choepel Y., Colangelo M.,
 RA Collins S., Collymore A., Cooke P., Deatellano K., Dewar K., Domino M.,
 RA Doyle M., Fensholt J., Ferrel P., Fitzhugh W., Forrest C., Gage D.,
 RA Galagan J., Gardyna S., Grant G., Hagos B., Heaford A., Horton L.,
 RA Howland J.C., Johnson R., Jones C., Kann L., Karatas A., Klein J.,
 RA Landers T., Lehoczeky J., Levine R., Liu C., Locke K.,
 RA MacDonald P., Margulis N., McEwan P., McGurk A., McKernan K., Mcpheeters R.,
 RA Macdonald P., Morrow J., Naylor J., Norman C.H., O'Connor T.,
 RA O'Donnell P., Oliver T.M., Peterson K., Pierre N., Pisanil C., Pollara V.,
 RA Raymond C., Riley R., Rothman D., Roy A., Santos R., Severy P., Spencer B.,
 RA Strange-Thoman N., Stojanovic N., Subramanian A., Talamas J., Testaye S.,
 RA Theodore J., Tirrell A., Vassiliev H., Viel R., Vo A., Wu X., Wyman D.,
 RA Ye W.J., Zimmer A., Zody M.;
 RT
 RT Submitted (31-DEC-1999) to the EMBL/GenBank/DBS databases.
 RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
 RL Cambridge, MA 02141, USA
 XX
 CC On Apr 1, 2000 this sequence version replaced gi:6721342.
 CC All repeats were identified using RepeatMasker:
 CC Smit, A.F.A. & Green, P. (1996-1997)
 CC http://ftp.genome.washington.edu/RM/RepeatMasker.html
 CC ----- Genome Center
 CC Center: Whitehead Institute/ MIT Center for Genome Research
 CC Center code: WIBR
 CC Web site: http://www-seq.wi.mit.edu
 CC Contact: sequence_submissions@genome.wi.mit.edu
 CC ----- Project Information
 CC Center project name: L3987
 CC -----
 CC Center clone name: 21_D_3
 CC ----- Summary Statistics
 CC Sequencing vector: M13; M77815; 100% of reads
 CC Chemistry: Dye-terminator Big Dye; 100% of reads
 CC Assembly program: Phrap; version 0.960731
 CC Consensus quality: 155438 bases at least Q40
 CC Consensus quality: 159063 bases at least Q20
 CC Consensus quality: 160301 bases at least Q20
 CC Insert size: 158000; agarose-fp
 CC Insert size: 161634; sum-of-ctrls
 CC Quality coverage: 4.8 in Q20 bases; agarose-fp
 CC Quality coverage: 4.7 in Q20 bases; sum-of-ctrls
 CC -----
 CC * NOTE: This is a 'working draft' sequence. It currently
 CC * consists of 16 contigs. The true order of the pieces
 CC * is not known and their order in this sequence record is
 CC * arbitrary. Gaps between the contigs are represented as
 CC * runs of N, but the exact sizes of the gaps are unknown.
 CC * This record will be updated with the finished sequence
 CC * as soon as it is available and the accession number will
 CC * be preserved.
 CC *
 CC 1 823: contig of 823 bp in length
 CC *
 CC 824 923: gap of 100 bp
 CC *
 CC 924 2980: contig of 2057 bp in length
 CC *
 CC 2981 3080: gap of 100 bp
 CC *
 CC 3081 5049: contig of 1965 bp in length
 CC *
 CC 5050 5149: gap of 100 bp
 CC *
 CC 5150 7713: contig of 2564 bp in length
 CC *
 CC 7714 7813: gap of 100 bp
 CC *
 CC 7814 10311: contig of 2498 bp in length
 CC *
 CC 10312 10411: gap of 100 bp
 CC *
 CC 10412 15992: contig of 5581 bp in length
 CC *
 CC 15993 16092: gap of 100 bp
 CC *
 CC 16093 22474: contig of 6382 bp in length
 CC *
 CC 22475 22574: gap of 100 bp
 CC *
 CC 22575 29644: contig of 7070 bp in length
 CC *
 CC 29645 29744: gap of 100 bp
 CC *
 CC 29745 38378: contig of 8634 bp in length
 CC *
 CC 38379 38478: gap of 100 bp
 CC *
 CC 38479 46868: contig of 8390 bp in length
 CC *
 CC 46869 46968: gap of 100 bp
 CC *
 CC 46969 56724: contig of 9756 bp in length

CC * 56725 56824: gap of 100 bp
 CC * 56825 66013: contig of 11189 bp in length
 CC * 66014 68113: gap of 100 bp
 CC * 68114 80779: contig of 12666 bp in length
 CC * 80780 80879: gap of 100 bp
 CC * 80880 94821: contig of 13942 bp in length
 CC * 94822 94921: gap of 100 bp
 CC * 94922 123346: contig of 28425 bp in length
 CC * 123347 123447: gap of 100 bp
 CC * 123447 163134: contig of 39688 bp in length.
 CC
 FH Key Location/Qualifiers
 FH
 FT source 1. 163134
 FT /db_xref="taxon:9606"
 FT /organism="Homo sapiens"
 FT /clone="RP11-21D3"
 FT /clone_lib="RP11 Human Male BAC"
 FT 1. 823
 FT /note="assembly-fragment-clone_end:T7-vector_side:right"
 FT 924. 2980
 FT /note="assembly-fragment"
 FT 3081. 5049
 FT /note="assembly-fragment"
 FT 5150. 7713
 FT /note="assembly-fragment"
 FT 7814. 10311
 FT /note="assembly-fragment"
 FT 10412. 15992
 FT /note="assembly-fragment"
 FT 16093. 22474
 FT /note="assembly-fragment"
 FT 22575. 29644
 FT /note="assembly-fragment"
 FT 29745. 38378
 FT /note="assembly-fragment"
 FT 38479. 46868
 FT /note="assembly-fragment"
 FT 46969. 56724
 FT /note="assembly-fragment"
 FT 56825. 68013
 FT /note="assembly-fragment"
 FT 68114. 80779
 FT /note="assembly-fragment"
 FT 80880. 94821
 FT /note="assembly-fragment"
 FT 94922. 123346
 FT /note="assembly-fragment-clone_end:SP6-vector_side:right"
 FT 123447. 163134
 FT /note="assembly-fragment"
 FT XX
 SQ Sequence 163134 BP; 43585 A; 37224 C; 37165 G; 43658 T; 1502 other;
 alignment_scores:
 alignment Quality: 9.00 Length: 9
 alignment Ratio: 1.000 Gaps: 0
 alignment Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-10-048-197-2 x AC019286 ..
 Align seq 1/1 to: AC019286 from: 1 to: 163134
 101 LysAsnLeuSerPheTyrLeuThrAla 109
 |||||
 149030 AAAAAGCTCAGTTCTACTTGACACGA 149056
 seq_name: gb_pr:AC108881
 seq_documentation_block: 163230 bp DNA linear PRI 01-FEB-2002
 LOCUS AC108881
 DEFINITION Homo sapiens chromosome X clone RP11-21D3 map Xp11.4, complete

sequence.
 AC108881 AC019286
 VERSION AC108881.1 GI:18464042
 KEYWORDS HTG: HTGS_DRAFT; HTGS_ACTIVEFIN.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 163230)
 Wen,G., Ramser,J., Baumgart,C., Melndl,A. and Platzner,M.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 163230)
 Wen,G. and Platzner,M.
 TITLE Direct Submission
 AUTHORS Submitted (01-FEB-2002) Genome Analysis, Institute of Molecular
 Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
 On Feb 1, 2002 this sequence version replaced accession g1:7382151.
 On Jan 22, 2002 this submission replaced accession AC019286
 Drifting Center: Whitehead Institute/MIT Center for Genome
 Research
 Flushing Center: Institute of Molecular Biotechnology
 Center: Institute of Molecular Biotechnology
 Center code: IMB
 Web site: http://genome.imb-jena.de/
 Contact: gscj-submit@genome.imb-jena.de
 Project Information
 Center project name: X154
 Center clone name: RP11-21D3
 Summary Statistics
 Sequencing vector: M13; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 158563 bases at least Q40
 Consensus quality: 160932 bases at least Q30
 Consensus quality: 161821 bases at least Q20
 Quality coverage: 5.31 x in Q20 bases; sum-of-contigs

Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality 10.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

FEATURES
 source 1..163230
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="Xp11.4"
 /clone="RP11-21D3"

BASE COUNT 45825 a 38652 c 36567 g 42186 t
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AC108881 ..

Align seg 1/1 to: AC108881 from: 1 to: 163230

101 LyAsnLeuSerPheTyrLeuThrAla 109
 |||||||
 124373 AAAAAGCTGAGTTCTACTGACAGCA 124399

seq_name: em_htg_hum:AC055844
 seq_documentation_block:
 ID AC055844 standard; DNA; HTG; 164017 BP.
 AC AC055844;
 XX AC055844.2
 SV AC055844.2
 XX 20-APR-2000 (Rel. 63, Created)
 DT 26-MAY-2000 (Rel. 63, Last updated, Version 2)
 XX 20-APR-2000 (Rel. 63, Last updated, Version 2)
 DE Homo sapiens chromosome 4 clone RP11-692K15 map 4, WORKING DRAFT SEQUENCE,
 DE 20 unordered pieces.
 DE
 XX HTG: HTGS_DRAFT; HTGS_PHASE1.
 XX
 XX Homo sapiens (human)
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC
 CC
 CC [1]
 RP Birren B., Linton L., Nusbaum C., Lander E.,
 RT "Homo sapiens chromosome 4, clone RP11-692K15";
 RL Unpublished.
 XX
 XX [2]
 RP 1-164017
 RP Birren B., Linton L., Nusbaum C., Lander E.,
 RA Anderson S., Baldwin J., Barna N., Bastien V., Bede F., Boguslavsky L.,
 RA Boukhalter B., Brown A., Burkett G., Campolano A., Castle A., Choepel Y.,
 RA Colangelo M., Collins S., Collymore A., Cooke P., Dearellano K., Dewar K.,
 RA Diaz J.S., Dodge S., Domino M., Doyle M., Ferreira P., Fitzhugh W.,
 RA Gage D., Galagan J., Gardyna S., Ginde S., Goyette M., Graham L.,
 RA Grand-Pierre N., Grant G., Hagos B., Heatford A., Horton L., Howland J.C.,
 RA Iliev I., Johnson R., Jones C., Kann L., Karatas A., Klein J., Larocque K.,
 RA Lamazares R., Landers T., Lehotzky J., Levine R., Liu C., Liu G.,
 RA Locke K., Macdonald P., Marquis N., McCarthy M., McEwan P., McGurk A.,
 RA McKernan K., McPheters R., Meldrum J., Menus L., Mihova T., Miranda C.,
 RA Mlenaga V., Morrow J., Murphy T., Naylor J., Norman C.H., O'Connor T.,
 RA O'Donnell P., O'Neill D., Olivari T.M., Oliver J., Peterson K., Pierre N.,
 RA Pisaní C., Pollara V., Raymond C., Riley R., Rogov P., Rothman D., Roy A.,
 RA Santos R., Schauer S., Severy P., Spencer B., Stange-Thomann N.,
 RA Stojanovic N., Sudramanian A., Talamas J., Testaye S., Theodore J.,
 RA Tirrell A., Travers M., Triggillo J., Vassiliev H., Viel R., Vo A.,
 RA Wilson B., Wu X., Wyman D., Ye W.J., Young G., Zalnoun J., Zimmer A.,
 RA Zody M.;
 RT
 RT Submitted (18-APR-2000) to the EMBL/GenBank/DBJ databases.
 RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
 RL Cambridge, MA 02141, USA
 RL
 XX
 CC On May 25, 2000 this sequence version replaced g1:7582677.
 CC All repeats were identified using RepeatMasker:
 CC Smit, A.F.A. & Green, P. (1996-1997)
 CC http://ftp.genome.washington.edu/RM/RepeatMasker.html
 CC
 CC ----- Genome Center
 CC Center: Whitehead Institute/ MIT Center for Genome Research
 CC Center code: MIBR
 CC Web site: http://www-seq.wi.mit.edu
 CC Contact: sequence_submissions@genome.wi.mit.edu
 CC
 CC ----- Project Information
 CC Center project name: L9704
 CC Center clone name: 692_K_15
 CC
 CC ----- Summary Statistics
 CC Sequencing vector: M13; M7815; 100% of reads
 CC Chemistry: Dye-terminator Big Dye; 100% of reads
 CC Assembly program: Phrap; version 0.960731
 CC Consensus quality: 153509 bases at least Q40
 CC Consensus quality: 159062 bases at least Q30
 CC Consensus quality: 161045 bases at least Q20
 CC Insert size: 170000; agarose-fp

```

CC      Insert size: 162117; sum-of-contigs
CC      Quality coverage: 4.2 in Q20 bases; agarose-fp
CC      Quality coverage: 4.4 in Q20 bases; sum-of-contigs
CC      -----
CC      * NOTE: This is a 'working draft' sequence. It currently
CC      * consists of 20 contigs. The true order of the pieces
CC      * is not known and their order in this sequence record is
CC      * arbitrary. Gaps between the contigs are represented as
CC      * runs of N, but the exact sizes of the gaps are unknown.
CC      * This record will be updated with the finished sequence
CC      * as soon as it is available and the accession number will
CC      * be preserved.
CC      *
CC      1 1095: contig of 1095 bp in length
CC      1096 1195: gap of 100 bp
CC      1196 3583: contig of 2388 bp in length
CC      3584 3683: gap of 100 bp
CC      3684 5501: contig of 1818 bp in length
CC      5502 5601: gap of 100 bp
CC      5602 7199: contig of 1598 bp in length
CC      7200 7299: gap of 100 bp
CC      7300 8463: contig of 1164 bp in length
CC      8464 8563: gap of 100 bp
CC      8564 12196: contig of 3633 bp in length
CC      12197 12296: gap of 100 bp
CC      12297 14633: contig of 2337 bp in length
CC      14634 14733: gap of 100 bp
CC      14734 17931: contig of 3198 bp in length
CC      17932 18031: gap of 100 bp
CC      18032 21385: contig of 3354 bp in length
CC      21386 21485: gap of 100 bp
CC      21486 25912: contig of 4427 bp in length
CC      25913 26012: gap of 100 bp
CC      26013 33022: contig of 7010 bp in length
CC      33023 33122: gap of 100 bp
CC      33123 40700: contig of 7578 bp in length
CC      40701 40800: gap of 100 bp
CC      40801 50958: contig of 10158 bp in length
CC      50959 51058: gap of 100 bp
CC      51059 58902: contig of 7844 bp in length
CC      58903 59002: gap of 100 bp
CC      59003 71011: contig of 12009 bp in length
CC      71012 71111: gap of 100 bp
CC      71112 81941: contig of 10830 bp in length
CC      81942 82041: gap of 100 bp
CC      82042 92606: contig of 10565 bp in length
CC      92607 92706: gap of 100 bp
CC      92707 105045: contig of 12339 bp in length
CC      105046 105145: gap of 100 bp
CC      105146 126729: contig of 21584 bp in length
CC      126730 126829: gap of 100 bp
CC      126830 164017: contig of 37188 bp in length.
CC      *
XX      Key      Location/Qualifiers
FH
FH      source      1. 164017
FT      /chromosome="4"
FT      /db_xref="taxon:9606"
FT      /organism="Homo sapiens"
FT      /map="4"
FT      /clone="RP11-692K15"
FT      /clone.lib="RP11 Human Male BAC"
FT      1. 1095
FT      misc_feature /note="assembly-fragment"
FT      1196. 3583
FT      misc_feature /note="assembly-fragment"
FT      3684. 5501
FT      misc_feature /note="assembly-fragment"
FT      5602. 7199
FT      misc_feature /note="assembly-fragment"
FT      7300. 8463
FT      misc_feature /note="assembly-fragment clone_end:SP6 vector_side:right"
FT      8564. 12196
FT      misc_feature /note="assembly-fragment"

```

```

FT      misc_feature 12297. 14633
FT      /note="assembly-fragment"
FT      misc_feature 14734. 17931
FT      /note="assembly-fragment"
FT      misc_feature 18032. 21385
FT      /note="assembly-fragment"
FT      misc_feature 21486. 25912
FT      /note="assembly-fragment"
FT      misc_feature 26013. 33022
FT      /note="assembly-fragment"
FT      misc_feature 33123. 40700
FT      /note="assembly-fragment"
FT      misc_feature 40801. 50958
FT      /note="assembly-fragment"
FT      misc_feature 51059. 58902
FT      /note="assembly-fragment"
FT      misc_feature 59003. 71011
FT      /note="assembly-fragment"
FT      misc_feature 71112. 81941
FT      /note="assembly-fragment"
FT      misc_feature 82042. 92606
FT      /note="assembly-fragment"
FT      misc_feature 92707. 105045
FT      /note="assembly-fragment"
FT      misc_feature 105146. 126729
FT      /note="assembly-fragment"
FT      misc_feature 126830. 164017
FT      /note="assembly-fragment clone_end:77 vector_side:right"
XX
SQ      Sequence 164017 BP; 51364 A; 29555 C; 30760 G; 50436 T; 1902 other;

```

```

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

```

```

alignment_block:
US-10-048-197-2 x AC055844 ..

```

```

Align seg 1/1 to: AC055844 from: 1 to: 164017

```

```

82 TrrpInguilleuSerleullele 90
|||||
45541 TGCACGAGATTTGCTTAATTATA 45567
seq_name: gb_hhg:AC010811

```

```

seq_documentation_block:
LOCUS      AC010811      165412 bp      DNA      linear      HTG 09-FEB-2002
DEFINITION Homo sapiens chromosome 11 clone RP11-3E17 map 11. *** SEQUENCING
IN PROGRESS ***; 5 unordered pieces.
ACCESSION  AC010811
VERSION     AC010811.10 GI:18642864
KEYWORDS   HTG; HTGS_PHASE1; HTGS_FUZZTOP; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 165412)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 11, clone RP11-3E17
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 165412)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Bana,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeRellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Garday,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karats,A., Klein,J.,
Lehoczky,J., Lien,C., Locke,K., Macdonald,P., Margulis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,

```

TITLE
JOURNAL

COMMENT
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Submitted (23-SEP-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 9, 2002 this sequence version replaced g1:18343112.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2720
Center clone name: 3_E_17

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 12794: contig of 12794 bp in length
* 12795 12894: gap of 100 bp
* 12895 68988: contig of 56094 bp in length
* 68989 69088: gap of 100 bp
* 69089 133991: contig of 64903 bp in length
* 133992 134091: gap of 100 bp
* 134092 162167: contig of 28076 bp in length
* 162168 162267: gap of 100 bp
* 162268 165412: contig of 3145 bp in length.
Location/Qualifiers
1. 165412
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-3P17"
/clone_lib="RP11-3P17 Human Male BAC"
BASE COUNT 48029 a 30561 c 34765 g 51633 t 424 others
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
Alignment_block:
us-10-048-197-2 x AC010811/rev ..
Align seg 1/1 to reverse of: AC010811 from: 1 to: 165412
85 TleuseerleulleiserPhelu 93
|||||
109495 ATCTCTTATATATCTCTTCTT 109469
seq_name: gp_hcg:AC106051
seq_documentation_block:
LOCUS AC106051 168577 bp DNA 11near HTG 11-JAN-2002
DEFINITION Homo sapiens chromosome 4 clone RP11-692K15, WORKING DRAFT
SEQUENCE, 3 unordered pieces.
ACCESSION AC106051 AC055844
VERSION AC106051.1 GI:18129600
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
human.
Homo sapiens
Eukaryota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 168577)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 168577)
Waterston, R.H.
Direct Submission
Submitted (11-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jan 11, 2002 this sequence version replaced g1:8084355.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@watson.wustl.edu
Center project name: H_NH0692K15
Drafting center: MIBR
----- Summary Statistics -----
Sequencing vector: M13, 528
Sequencing vector: plasmid, 488
Chemistry: Dye-terminator Big Dye, 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 167893 bases at least Q40
Consensus quality: 168622 bases at least Q30
Insert size: 17000; agarose-fp
Insert size: 172425; sum-of-contigs
Quality coverage: 9.25 in Q20 bases; agarose-fp
Quality coverage: 8.49 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1210: contig of 1210 bp in length
* 1211 1310: gap of unknown length
* 1311 35898: contig of 34588 bp in length
* 35899 168577: gap of unknown length
* 168578 168577: contig of 13579 bp in length.
Location/Qualifiers
1. 168577
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-692K15"
/note="assembly_name:Contig17"
1311. 35898
/note="assembly_name:Contig18"
35899. 168577
/note="assembly_name:Contig19
clone_end:17
vector_side:right"
BASE COUNT 53348 a 31187 c 31329 g 52513 t 200 others
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AC106051 ..

Align seg 1/1 to: AC106051 from: 1 to: 168577

82 TrpGlnGluLeuSerLeuLeu 90
|||||
70847 TGCAGAGATTTCTTTAATTAATA 70873

seq_name: gb-pr:AC097661

seq_documentation_block:

LOCUS AC097661 169197 bp DNA linear PRI 03-JAN-2002
DEFINITION Homo sapiens chromosome 4 clone RP11-553E15, complete sequence.
ACCESSION AC097661 AC023845
VERSION AC097661.3 GI:17975439
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 169197)
TITLE The sequence of Homo sapiens clone

JOURNAL Unpublished
2 (bases 1 to 169197)

REFERENCE Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 3 (bases 1 to 169197)
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 169197)
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE On Dec 21, 2001 this sequence version replaced gi:17298655.
COMMENT

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
Project Information
Center project name: H.NH0353E15
Drafting center: WIBR

FEATURES
source
1. 169197
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-553E15"
BASE COUNT 56173 a 31106 c 29512 g 52406 t
ORIGIN

Alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AC097661/rev ..

Align seg 1/1 to reverse of: AC097661 from: 1 to: 169197

10 HisSerAlaLeuThrAlaLeu 18
|||||
126364 ATTTCTGCCCCCTGCTACAGCTTACTC 126388

seq_name: gb-ro:AC074041

seq_documentation_block:

LOCUS AC074041 170548 bp DNA linear ROD 28-JUN-2001
DEFINITION Mus musculus chromosome 2 clone RP23-258D16 strain C57BL6/J, complete sequence.
ACCESSION AC074041
VERSION AC074041.3 GI:14573692
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 170548)
TITLE Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Granite, S., Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E., Lee-Jin, S.-Q., Legaspi, R., Lim, M., Maduro, Q.L., Maduro, V.B., Mastello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y., Snyder, B., Stantir, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgren, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Zhang, L.-H. and Green, E.D.
TITLE NISC Comparative Sequencing Initiative
JOURNAL Unpublished
2 (bases 1 to 170548)

REFERENCE Green, E.D.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2000) NIH Intramural Sequencing Center, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 170548)

REFERENCE Green, E.D.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-2001) NIH Intramural Sequencing Center, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA
On Jun 28, 2001 this sequence version replaced gi:11024833.

COMMENT
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nigrl.nih.gov
Project Information
Center project name: ta
Center clone name: 258D16

This sequence was finished as follows unless otherwise noted:
all regions were double-stranded, sequenced with an
alternate chemistry, or covered by high quality data
(i.e., phred quality >= 30); an attempt was made to resolve
all sequencing problems, such as compressions and repeats;
or more than one M13 subclone; and the assembly was confirmed
by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of
this clone unless otherwise noted. If there are overlapping
clones, the overlaps are noted in the beginning and end of
the features section.

FEATURES
source
1. 170548
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-258D16"
/clone_lib="RPCI mouse BMC library 23"
misc_feature 1. 17547

/note="clone overlaps with GenBank Accession Number
AC074224 (nucleotides 171443-188989) clone RP23-43C3
(center project name uz)"

misc_feature

170267..170274
/note="low quality single stranded/single chemistry
region"

BASE COUNT 51882 a 37418 c 35891 g 45357 t
ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AC074041/rev ..

Align seg 1/1 to reverse of: AC074041 from: 1 to: 170548

8 leuThrlleSerAlaIleuIeuThraA 16

|||||
9603 TTGACCAATTGCGGTGTAAACGGCT 9577

seq_name: gb_htg:AC099145

seq_documentation_block:

LOCUS AC099145 172176 bp DNA linear HTG 20-DEC-2001

DEFINITION Rattus norvegicus clone CH230-68A17, *** SEQUENCING IN PROGRESS

***, 75 unordered pieces.

ACCESSION AC099145

VERSION AC099145.2 GI:17973022

KEYWORDS HTG: HTGS_PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 172176)

Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
Alstodts,S.L., Amartunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blimie,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brilewa,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dlnh,H.H., Douhaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Gatta,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hayes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichteage,O., Lieu,C., Liu,J., Liu,W.,
Lounsged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mayhew,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabdat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,N., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenko,S.,
Ogulu,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshart,N.,
Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabar,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleciyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 172176)

Worley,K.C.

Direct Submission

Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:16874639.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GGYO

Center clone name: CH230-68A17

Summary Statistics

Assembly program: Phrap; version 0.990329first call to

findhaplolist

Consensus quality: 137202 bases at least Q40

Consensus quality: 145881 bases at least Q30

Consensus quality: 153638 bases at least Q20

Estimated insert size: 146753; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 1.9x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	5563: contig of 5563 bp in length
5564	5663: gap of unknown length
5664	11484: contig of 5821 bp in length
11485	11584: gap of unknown length
11585	14678: contig of 3094 bp in length
14679	14778: gap of unknown length
14779	18800: contig of 4022 bp in length
18801	18900: gap of unknown length
18901	23194: contig of 4294 bp in length
23195	23294: gap of unknown length
23295	26458: contig of 3164 bp in length
26459	26558: gap of unknown length
26559	29915: contig of 3357 bp in length
29916	30015: gap of unknown length
30016	34394: contig of 4379 bp in length
34395	34494: gap of unknown length
34495	37302: contig of 2708 bp in length
37303	37302: gap of unknown length
37303	41893: contig of 4591 bp in length
41894	41993: gap of unknown length
41994	44551: contig of 2558 bp in length
44552	44651: gap of unknown length
44652	46827: contig of 2176 bp in length
46828	46927: gap of unknown length
46928	49389: contig of 2462 bp in length
49389	49489: gap of unknown length
49490	52444: contig of 2955 bp in length
52445	52544: gap of unknown length
52545	55673: contig of 3129 bp in length
55674	55773: gap of unknown length
55774	58799: contig of 3026 bp in length

```

* 58800 58899: gap of unknown length
* 58900 60728: contig of 1829 bp in length
* 60729 60828: gap of unknown length
* 60829 64135: contig of 3307 bp in length
* 64136 64235: gap of unknown length
* 64236 66744: contig of 2509 bp in length
* 66745 66844: gap of unknown length
* 66845 70140: contig of 3296 bp in length
* 70141 70240: gap of unknown length
* 70241 72453: contig of 2213 bp in length
* 72454 72553: gap of unknown length
* 72554 75923: contig of 3370 bp in length
* 75924 76023: gap of unknown length
* 76024 77534: contig of 1411 bp in length
* 77535 80128: contig of 2593 bp in length
* 80129 80227: gap of unknown length
* 80228 82231: contig of 2004 bp in length
* 82232 82331: gap of unknown length
* 82332 83813: contig of 1482 bp in length
* 83814 83913: gap of unknown length
* 83914 86543: contig of 2630 bp in length
* 86544 87958: gap of unknown length
* 87959 88059: contig of 1316 bp in length
* 88060 91075: gap of unknown length
* 91076 91175: contig of 3016 bp in length
* 91176 92899: contig of 1724 bp in length
* 92900 95452: gap of unknown length
* 95453 95553: contig of 2453 bp in length
* 95554 96932: gap of unknown length
* 96933 97032: contig of 1380 bp in length
* 97033 99660: gap of unknown length
* 99661 99760: contig of 2628 bp in length
* 99761 101461: gap of unknown length
* 101462 101561: contig of 1701 bp in length
* 101562 103063: gap of unknown length
* 103064 103163: contig of 1502 bp in length
* 103164 104494: gap of unknown length
* 104495 104594: contig of 1331 bp in length
* 104595 107379: gap of unknown length
* 107380 107479: contig of 2785 bp in length
* 107480 110000: gap of unknown length
* 110001 110100: contig of 2521 bp in length
* 110101 112185: gap of unknown length
* 112186 112285: contig of 2085 bp in length
* 112286 114526: gap of unknown length
* 114527 114626: contig of 2241 bp in length
* 114627 116967: gap of unknown length
* 116968 117067: contig of 2341 bp in length
* 117068 118286: gap of unknown length
* 118287 118386: contig of 1219 bp in length
* 118387 119705: gap of unknown length
* 119706 119805: contig of 1319 bp in length
* 119806 120897: gap of unknown length
* 120898 120997: contig of 1092 bp in length
* 120998 123513: gap of unknown length
* 123514 123614: contig of 2516 bp in length
* 123615 125484: gap of unknown length
* 125485 125584: contig of 1871 bp in length
* 125585 128208: gap of unknown length
* 128209 128308: contig of 2624 bp in length
* 128309 129589: gap of unknown length
* 129590 129689: contig of 1281 bp in length
* 129690 131532: gap of unknown length
* 131533 131632: contig of 1843 bp in length
* 131633 133157: gap of unknown length
* 133158 133257: contig of 1525 bp in length
* 133258 135339: gap of unknown length
* 135339 135639: contig of 2282 bp in length
* 135640 136898: gap of unknown length
* 136899 136998: contig of 1259 bp in length

```

```

* 136999 138251: contig of 1253 bp in length
* 138252 138351: gap of unknown length
* 138352 139865: contig of 1514 bp in length
* 139866 139966: gap of unknown length
* 139967 141534: contig of 1569 bp in length
* 141535 141635: gap of unknown length
* 141636 143556: contig of 1922 bp in length
* 143557 143656: gap of unknown length

```

alignment_scores:

```

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

```

US-10-048-197-2 x AC099145 ..

```

```

Align seg 1/1 to: AC099145 from: 1 to: 172176

```

```

47 Thraanlysgingluileserserarg 55
|||||

```

```

126728 ACCAATAACAGAAATCTCAGACAGA 126754

```

```

seq_name: gb_htg:AC013798

```

```

seq_documentation_block:

```

```

LOCUS AC013798 175706 bp DNA linear HTG 01-APR-2000
DEFINITION Homo sapiens clone RP11-20114, WORKING DRAFT SEQUENCE, 13 unordered
pieces.

```

```

ACCESSION AC013798
VERSION AC013798.4 GI:7382144
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

```

```

SOURCE human.

```

```

ORGANISM Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

1 (bases 1 to 175706)
Birten,B., Linton,L., Nusbaum,C. and Lande,E.

```

```

Homo sapiens chromosome, clone RP11-20114
unpublished

```

```

2 (bases 1 to 175706)
Birten,B., Linton,L., Nusbaum,C., Lande,E., Allen,N., Anderson,M.,

```

```

Baldwin,J., Barina,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,

```

```

Cooke,P., Darrellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,

```

```

Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,K., Klein,J.,

```

```

Lehoczky,J., Lien,C., Locke,K., Macdonald,P., Margulis,N.,
McEwan,P., McGuirk,A., McKernan,K., McDonald,L., Marquis,N.,

```

```

Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,

```

```

Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessie,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,

```

```

Wymann,D., Ye,W.J., Zimmer,A. and Zody,M.

```

```

Direct Submissions

```

```

Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

```

On Apr 1, 2000 this sequence version replaced gi:554527.

```

```

All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)

```

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

Genome Center

```

```

Center: Whitehead Institute/ MIT Center for Genome Research

```

```

Center code: WIRB

```

```

Web site: http://www-seq.wi.mit.edu

```

```

Contact: sequence_submissions@genome.wi.mit.edu

```

```

Project Information

```

```

Center project name: I3908

```

```

Center clone name: 20.I.14

```

```

Summary Statistics

```

```

Sequencing vector: M13; M7815; 100% of reads

```

```

Chemistry: Dye-terminator Big Dye; 100% of reads

```

```

Assembly program: tmap; version 0.960731
Consensus quality: 166346 bases at least Q40
Consensus quality: 169904 bases at least Q30
Consensus quality: 171697 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 174506; sum-of-contigs
Quality coverage: 5.2 in Q20 bases; agarose-fp
Quality coverage: 5.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1504: contig of 1504 bp in length
* 1505 1604: gap of 100 bp
* 1605 4186: contig of 2582 bp in length
* 4187 4286: gap of 100 bp
* 4287 6724: contig of 2438 bp in length
* 6725 6824: gap of 100 bp
* 6825 13163: contig of 6339 bp in length
* 13164 13263: gap of 100 bp
* 13264 18433: contig of 5170 bp in length
* 18434 18533: gap of 100 bp
* 18534 27908: contig of 9375 bp in length
* 27909 28008: gap of 100 bp
* 28009 34485: contig of 6677 bp in length
* 34486 34585: gap of 100 bp
* 34586 44981: contig of 10396 bp in length
* 44982 45081: gap of 100 bp
* 45082 56951: contig of 11870 bp in length
* 56952 57051: gap of 100 bp
* 57052 67256: contig of 10205 bp in length
* 67257 67356: gap of 100 bp
* 67357 97219: contig of 29863 bp in length
* 97220 97319: gap of 100 bp
* 97320 134359: contig of 37040 bp in length
* 134360 134459: gap of 100 bp
* 134460 175706: contig of 41247 bp in length.
*
* Location/Qualifiers
* 1..175706
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /clone_11b="PC11-20114"
* /clone_11b="PC11 Human Male BAC"
* 1..1504
* /note="assembly_fragment"
* 1605..4186
* /note="assembly_fragment"
* 4287..6724
* /note="assembly_fragment"
* 6825..13163
* /note="assembly_fragment"
* 13264..18433
* /note="assembly_fragment"
* clone_end:17
* vector_side:left"
* 18534..27908
* /note="assembly_fragment"
* 28009..34485
* /note="assembly_fragment"
* 34586..44981
* /note="assembly_fragment"
* 45082..56951
* /note="assembly_fragment"
* 57052..67256
* /note="assembly_fragment"
* clone_end:SP6
* vector_side:left"
* 67357..97219

```

```

misc_feature          /note="assembly_fragment"
97320..134359        /note="assembly_fragment"
misc_feature          /note="assembly_fragment"
134460..175706       /note="assembly_fragment"
BASE COUNT           46950 a 40425 c 40682 g 46442 t 1207 others
ORIGIN

alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AC013798 ..

Align seg 1/1 to: AC013798 from: 1 to: 175706

101 LysAsnLeuSerPheTyrLeuThra1a 109
|||||
9798 AAAAACCTGATTTCTACTTGACAGCA 9824

seq_name: gb_hgt.AC009290

seq_documentation_block:
LOCUS              178670 bp      DNA      linear      HTG 04-MAY-2001
DEFINITION        Homo sapiens clone RP11-45118, WORKING DRAFT SEQUENCE, 10 unordered
                    pieces.
ACCESSION          AC009290
VERSION            AC009290.3 GI:8072429
KEYWORDS           HTG; PHASE1; HTGS_DRAFT.
SOURCE             human.
ORGANISM           Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE          1 (bases 1 to 178670)
AUTHORS            Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
TITLE              Baker,J., Baldwin,J., Barna,N., Beckley,R., Benn,J., Brown,A.,
JOURNAL            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A.,
REFERENCE          Cooke,P., Dettellano,K., Depayre,E., Devon,K., Dewar,K.,
AUTHORS            Donehan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
TITLE              Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
JOURNAL            Hatos,B., Heatford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
REFERENCE          Karas,A., Lehoczy,J., Lieu,C., Locke,K., MacDonald,P.,
AUTHORS            Marquis,N., McEwan,P., McGuirk,A., McKernan,K., McLaughlin,J.,
TITLE              Melidrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
JOURNAL            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
REFERENCE          Peterson,K., Poltara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
AUTHORS            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
TITLE              Tesfaye,S., Toruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
JOURNAL            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
COMMENT            Direct Submission
                    Submitted (13-AUG-1999) Whitehead Institute/MIT Center for Genome
                    Research, 320 Charles Street, Cambridge, MA 02141, USA
                    On May 25, 2000 this sequence version replaced gi:7328737.
                    All repeats were identified using RepeatMasker:
                    Smit, A.F.A. & Green, P. (1996-1997)
                    http://ftp.genome.washington.edu/RM/RepeatMasker.html
                    ----- Genome Center
                    Center: Whitehead Institute/ MIT Center for Genome Research
                    Center code: WIBR
                    Web site: http://www-seq.wi.mit.edu
                    Contact: sequence_submissions@genome.wi.mit.edu
                    ----- Project Information
                    Center project name: L1165
                    Center clone name: 45_I18
                    ----- Summary Statistics
                    Sequencing vector: M13; M77815; 93% of reads

```


Sequencing vector: Plasmid: n/a; %0.1% of reads
 7.38747553816047Chemistry: Dye primer: amersham; 8% of reads
 Chemistry: Dye terminator: Big Dye; 92% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 171189 bases at least Q40
 Consensus quality: 174658 bases at least Q30
 Consensus quality: 176319 bases at least Q20
 Insert size: 190000; agarose-gel
 Insert size: 177770; sum-of-contigs

Quality coverage:
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1183: contig of 1183 bp in length
 1184 1283: gap of 100 bp
 1284 3307: contig of 2024 bp in length
 3308 3407: gap of 100 bp
 3408 7626: contig of 4218 bp in length
 7627 7726: gap of 100 bp
 7727 15657: contig of 8931 bp in length
 15658 16757: gap of 100 bp
 16758 26995: contig of 10238 bp in length
 26996 27095: gap of 100 bp
 27096 42276: contig of 15181 bp in length
 42277 42376: gap of 100 bp
 42377 60953: contig of 18577 bp in length
 60954 61053: gap of 100 bp
 61054 83707: contig of 22654 bp in length
 83708 83807: gap of 100 bp
 83808 116924: contig of 33117 bp in length
 116925 117024: gap of 100 bp
 117025 178670: contig of 61646 bp in length.

FEATURES

source
 1. 178670
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="RP11-45118"
 /clone_lib="RP11-45118 Human Male BAC"
 1. 1183
 /note="assembly-fragment"
 misc_feature
 1284..3307
 /note="assembly-fragment"
 misc_feature
 3408..7626
 /note="assembly-fragment"
 misc_feature
 7727..16657
 /note="assembly-fragment"
 misc_feature
 16758..26995
 /note="assembly-fragment"
 misc_feature
 27096..42276
 /note="assembly-fragment"
 misc_feature
 42377..60953
 /note="assembly-fragment"
 clone_end:r7
 vector_side:left
 61054..83707
 /note="assembly-fragment"
 misc_feature
 83808..116924
 /note="assembly-fragment"
 clone_end:SP6
 vector_side:left
 117025..178670
 /note="assembly-fragment"
 BASE COUNT 57920 a 31871 c 32262 g 55715 t 902 others
 ORIGIN

alignment_scores:
 Quality: 9.00

Length: 9

Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:

US-10-048-197-2 x AC009290/rev ..

Align seg 1/1 to reverse of: AC009290 from: 1 to: 178670

86 Leuserleu1le1serphelutrip 94

59489 CTTTCATTAATATATATATATATATATATG 59463

seq_name: gb.ro:AC074224

seq_documentation block:

LOCUS AC074224 188989 bp DNA linear ROD 22-JAN-2001

DEFINITION Mus musculus chromosome 2 clone RP23-43c3 strain C57BL/6/J, complete

sequence.

ACCESSION AC074224

VERSION AC074224.3 GI:12331469

KEYWORDS HTG.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE

Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,

Boulford, G.G., Brinley, C., Brooks, S., Dietrich, N.L., Granite, S.,

Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E., Lee-Ilh, S.-Q.,

Legaspi, R., Lim, M., Maduro, O.L., Maduro, V.B., Mastaglio, C.,

Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A.,

Shenchenko, Y., Snyder, B., Stantipop, S., Thomas, J.W., Thomas, P.J.,

Tingson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,

Wecherdy, K.D., Zhang, L.-H. and Green, E.D.

NISC Comparative Sequencing Initiative

Unpublished

2 (bases 1 to 188989)

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

TITLE

CLONE LENGTH: This sequence represents the entire insert of
 this clone unless otherwise noted. If there are overlapping
 clones, the overlaps are noted in the beginning and end of
 the features section.

FEATURES

source
 1. 188989
 /organism="Mus musculus"
 /strain="C57BL6/J"
 /db_xref="taxon:10090"


```

/chromosome="2"
/clone_1lb="RP23-43C3"
misc_feature 75707..75760 /note="pcr product sequence only"
misc_feature 83183..83273 /note="single clone coverage"
misc_feature 103600..103772 /note="single clone coverage"
BASE COUNT 53467 a 42113 c 41895 g 51514 t
ORIGIN

alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

Alignment block:
US-10-048-197-2 x AC074224/rev ..

Align seg 1/1 to reverse of: AC074224 from: 1 to: 188989
      8 LeuThrilleseraLeuLeuThrala 16
      |||||
181045 TTGACCATTCGGCGTGTACGCGCT 181019

seq_name: gb_hvg:AC023140

seq_documentation_block:
LOCUS AC023140 194575 bp DNA linear HVG 07-JUL-2000
DEFINITION Homo sapiens chromosome 15 clone RP11-535P8, WORKING DRAFT
ACCESSION AC023140
VERSION AC023140.4 GI:8570285
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 194575)
REFERENCE 1 (bases 1 to 194575)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 194575)
REFERENCE Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2000) Genome Sequencing Center, Washington
AUTHORS University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA
COMMENT On Jun 17, 2000 this sequence version replaced gl:7232193.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0535P08
----- Summary Statistics -----
Sequencing vector: M13: 1008
Sequencing vector: Plasmid: 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 183084 bases at least Q40
Consensus quality: 18527 bases at least Q40
Consensus quality: 188372 bases at least Q20
Insert size: 209000; agarose-fp
Insert size: 192075; sum-of-contigs
Quality coverage: 3.93 in Q20 bases; agarose-fp
Quality coverage: 4.32 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently

```

```

* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1295: contig of 1295 bp in length
1296 1395: gap of unknown length
1396 3355: contig of 1960 bp in length
3355 3356: gap of unknown length
3356 5113: contig of 1658 bp in length
5113 5214: gap of unknown length
5214 6910: contig of 1697 bp in length
6910 7010: gap of unknown length
7010 9042: contig of 2032 bp in length
9042 9043: gap of unknown length
9043 9142: contig of 2503 bp in length
9142 11645: contig of 2503 bp in length
11645 11745: gap of unknown length
11745 14544: contig of 2799 bp in length
14544 14545: gap of unknown length
14545 17322: contig of 2678 bp in length
17322 17422: gap of unknown length
17422 19837: contig of 2415 bp in length
19837 19937: gap of unknown length
19937 25695: contig of 5758 bp in length
25695 25795: gap of unknown length
25795 29231: contig of 3436 bp in length
29231 29332: gap of unknown length
29332 32892: contig of 3561 bp in length
32892 32993: gap of unknown length
32993 38263: contig of 5271 bp in length
38263 38264: gap of unknown length
38264 42404: contig of 4041 bp in length
42404 42504: gap of unknown length
42504 48404: contig of 5900 bp in length
48404 48504: gap of unknown length
48504 57280: contig of 8786 bp in length
57280 57291: gap of unknown length
57291 57391: contig of 8513 bp in length
57391 65905: contig of 8513 bp in length
65905 66006: gap of unknown length
66006 74728: contig of 8723 bp in length
74728 74729: gap of unknown length
74729 87289: contig of 12461 bp in length
87289 87390: gap of unknown length
87390 99467: gap of 12078 bp in length
99467 99567: gap of unknown length
99567 111113: contig of 11546 bp in length
111113 11213: gap of unknown length
11213 122406: contig of 11193 bp in length
122406 122407: gap of unknown length
122407 122507: gap of 12974 bp in length
122507 135481: contig of 12974 bp in length
135481 135581: gap of unknown length
135581 153618: contig of 18038 bp in length
153618 153718: gap of unknown length
153718 170017: contig of 16299 bp in length
170017 170117: gap of unknown length
170117 194575: contig of 24458 bp in length.
Location/Qualifiers
1. 194575
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/clone="RP11-535P8"
1. 1295
/note="assembly_name:Contig9"
misc_feature 1396..3355
/note="assembly_name:Contig10"
misc_feature 3456..5113
/note="assembly_name:Contig11"
5214..6910
/note="assembly_name:Contig12"
7011..9042
misc_feature

```

```

misc_feature /note="assembly_name:Contig13"
9143..11645
misc_feature /note="assembly_name:Contig14"
11746..14544
misc_feature /note="assembly_name:Contig15"
14645..17322
/note="assembly_name:Contig16
clone_end:17
vector_side:left"
misc_feature /note="assembly_name:Contig17"
17423..19837
misc_feature /note="assembly_name:Contig18"
19938..225695
misc_feature /note="assembly_name:Contig19"
225796..29231
misc_feature /note="assembly_name:Contig20"
29332..32892
misc_feature /note="assembly_name:Contig21"
32993..38263
/note="assembly_name:Contig22"
38364..42404
misc_feature /note="assembly_name:Contig23"
42505..48404
/note="assembly_name:Contig24"
48505..57290
misc_feature /note="assembly_name:Contig25"
57391..65905
/note="assembly_name:Contig26"
66006..74728
misc_feature /note="assembly_name:Contig27"
74829..87289
/note="assembly_name:Contig28"
87390..99467
misc_feature /note="assembly_name:Contig29"
99568..111113
misc_feature /note="assembly_name:Contig30"
11214..122406
/note="assembly_name:Contig31"
122507..135480
misc_feature /note="assembly_name:Contig32"
135581..153618
/note="assembly_name:Contig33"
153719..170017
misc_feature /note="assembly_name:Contig34"
170118..194575
vector_side:right"
misc_feature /note="assembly_name:Contig34"
170118..194575
BASE COUNT 58990 a 40026 c 37747 g 55237 t 2575 others
ORIGIN
alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x AC023140/rev ..
Align seg 1/1 to reverse of: AC023140 from: 1 to: 194575
85 l1e1euserleutle1e1euserPhetleu 93
|||||
73566 ATTCTCTCTCTATTCATTCCTTCCTG 73540
seq_name: gb_h1tg:AC068980
seq_documentation_block:
LOCUS AC068980 204991 bp DNA linear HTG 05-OCT-2001
DEFINITION Homo sapiens chromosome 3 clone RP11-352A14, WORKING DRAFT
SEQUENCE 8 unordered pieces.
ACCESSION AC068980
VERSION AC068980.10 GI:15741365

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

HTG; HTGS_PHASE1; HTGS_DRAFT.
human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 204991)

Muzny,D.M., Adams,C., Adl-Oudula,B., All-ouman,F.R., Allen,C.,
Alstbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blum,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,T.,
Foster,P., Frank,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,M., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,D., Liu,W.,
Luiseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., Meleod,M.P., Meador,M.,
Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,S., Nguyen,N., Nickerson,E., Nockenwo,S.,
Ogih,M., Okwouu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Plums,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshbari,N.,
Slisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleciyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

COMMENT

Submitted (15-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 25, 2001 this sequence version replaced gi:11128156.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center Project name: HB9Y

Center clone name: RP11-552A14

Summary Statistics

Sequencing vector: M13; L08821

Chemistry: Dye-Primer Bodypy; 14% of reads

Chemistry: Dye-terminator Big Dye; 86% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 196487 bases at least Q40

Consensus quality: 203036 bases at least Q30

Consensus quality: 206264 bases at least Q20

Estimated insert size: 205706; sum-of-coverage estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 4.6x in Q20 bases; sum-of-coverage estimation

```

*****
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draftc.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 52522: contig of 52522 bp in length
* 52523: gap of unknown length
* 52623: contig of 38737 bp in length
* 91360: gap of unknown length
* 91459: gap of unknown length
* 125166: contig of 33707 bp in length
* 125167: gap of unknown length
* 125267: contig of 27631 bp in length
* 152898: gap of unknown length
* 152997: gap of unknown length
* 152998: contig of 20604 bp in length
* 173601: gap of unknown length
* 173702: contig of 14946 bp in length
* 186647: gap of unknown length
* 186747: gap of unknown length
* 188648: gap of unknown length
* 202570: contig of 13823 bp in length
* 202571: gap of unknown length
* 202670: gap of unknown length
* 202671: gap of unknown length
* 204991: contig of 2321 bp in length.
*
* Location/Qualifiers
*
*   .204991
*   /organism="Homo sapiens"
*   /db_xref="taxon:9606"
*   /chromosome="3"
*   /clone="RP11-552A14"
*
BASE COUNT    60973 a 40617 c 40715 g 61970 t    716 others
ORIGIN

```

```

*****
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1059: contig of 1059 bp in length
* 1060: gap of 100 bp
* 1160: 3098: contig of 1939 bp in length
* 3099: 3198: gap of 100 bp
* 3199: 4151: contig of 953 bp in length
* 4152: 4251: gap of 100 bp
* 4252: 6077: contig of 1826 bp in length
* 6078: 6177: gap of 100 bp
* 6178: 10158: contig of 3981 bp in length
* 10159: 10258: gap of 100 bp
* 10259: 13440: contig of 3182 bp in length
* 13441: 13540: gap of 100 bp
* 13541: 16429: contig of 2889 bp in length
* 16430: 16529: gap of 100 bp
* 16530: 19606: contig of 3077 bp in length
* 19607: 19706: gap of 100 bp
* 19707: 24259: contig of 4553 bp in length
* 24260: 24359: gap of 100 bp
* 24360: 28314: contig of 3955 bp in length
* 28315: 28414: gap of 100 bp
* 28415: 32584: contig of 4170 bp in length
* 32585: 32684: gap of 100 bp
* 32685: 37793: contig of 5109 bp in length
* 37794: 37893: gap of 100 bp

```

```

* 37894 44505: contig of 6612 bp in length
* 44506 44605: gap of 100 bp
* 44606 50170: contig of 5565 bp in length
* 50171 50270: gap of 100 bp
* 50271 57312: contig of 6942 bp in length
* 57313 57312: gap of 100 bp
* 57313 68003: contig of 10691 bp in length
* 68004 68103: gap of 100 bp
* 68104 77484: contig of 9381 bp in length
* 77485 77584: gap of 100 bp
* 77585 86863: contig of 9279 bp in length
* 86864 86963: gap of 100 bp
* 86964 96495: contig of 9532 bp in length
* 96496 96595: gap of 100 bp
* 96596 107929: contig of 11334 bp in length
* 107930 108029: gap of 100 bp
* 108030 116856: contig of 8827 bp in length
* 116857 116956: gap of 100 bp
* 116957 128225: contig of 11269 bp in length
* 128226 128325: gap of 100 bp
* 128326 142566: contig of 14241 bp in length
* 142567 142666: gap of 100 bp
* 142667 160140: contig of 17474 bp in length
* 160141 160240: gap of 100 bp
* 160241 179301: contig of 19061 bp in length
* 179302 179401: gap of 100 bp
* 179402 204993: contig of 25592 bp in length.
FEATURES
    source
        1. 204993
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="3"
            /map="3"
            /clone="RP11-479J2"
            /clone_lib="RP11 Human Male BAC"
            1. .1059
                /note="assembly-fragment"
                1160. .3098
                    /note="assembly-fragment"
                    3199. .4151
                        /note="assembly-fragment"
                        clone_end:r7
                        vector_side:left"
                        4252. .6077
                            /note="assembly-fragment"
                            6178. .10158
                                /note="assembly-fragment"
                                10259. .13440
                                    /note="assembly-fragment"
                                    13441. .16429
                                        /note="assembly-fragment"
                                        16530. .19606
                                            /note="assembly-fragment"
                                            19707. .24259
                                                /note="assembly-fragment"
                                                24360. .28314
                                                    /note="assembly-fragment"
                                                    28415. .32584
                                                        /note="assembly-fragment"
                                                        32685. .37793
                                                            /note="assembly-fragment"
                                                            37894. .44505
                                                                /note="assembly-fragment"
                                                                44606. .50170
                                                                    /note="assembly-fragment"
                                                                    50271. .57212
                                                                        /note="assembly-fragment"
                                                                        57313. .68003
                                                                            /note="assembly-fragment"
                                                                            68104. .77484
                                                                                /note="assembly-fragment"
                                                                                77585. .86863
                                                                                    /note="assembly-fragment"

```

```

misc_feature 86964. .96495
              /note="assembly-fragment"
misc_feature 96596. .107929
              /note="assembly-fragment"
misc_feature 108030. .116856
              /note="assembly-fragment"
              clone_end:SP6
              vector_side:left"
misc_feature 116957. .128225
              /note="assembly-fragment"
misc_feature 128326. .142566
              /note="assembly-fragment"
misc_feature 142667. .160140
              /note="assembly-fragment"
misc_feature 160241. .179301
              /note="assembly-fragment"
misc_feature 179402. .204993
              /note="assembly-fragment"
BASE COUNT 62968 a 39652 c 40375 g 59489 t 2509 others
ORIGIN

```

```

alignment_scores:
    Quality: 9.00
    Ratio: 1.000
    Percent Similarity: 100.000
    Gaps: 0
    Percent Identity: 100.000

```

```

alignment_block:
    US-10-048-197-2 x AC024447
    ..

```

```

Align seg 1/1 to: AC024447 from: 1 to: 204993

```

```

8 Leuthri1esR1aleuLeuthria 16
|||||
180405 CTCACAACTCTGCTCTGACGACGA 180431

```

```

seq_name: gb_htg:AC092942

```

```

seq_documentation_block:
    LOCUS AC092942 205825 bp DNA linear HTG 03-JAN-2002
    DEFINITION Homo sapiens chromosome 3q clone RP11-479J2, WORKING DRAFT
    ACCESSION AC092942
    VERSION AC092942.2 GI:18034707
    KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
    SOURCE human.
    ORGANISM Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
        1 (bases 1 to 205825)
        Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
        Albrooks,S.L., Amaralung,H.C., Are,J.R., Banks,T., Barberia,J.,
        Benton,J., Bimaga,K., Blankenburg,K., Bonini,D., Bouck,J.,
        Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhy,C.,
        Butcher,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
        Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
        Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
        Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
        Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
        Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
        Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
        Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
        Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
        Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
        Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
        Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
        Hollins,B., Honsi,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
        Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
        Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,D.,
        Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
        Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
        Louised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
        Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

```

Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, N., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkwo, S., Ogih, M., Okunolu, G., Oragunye, N., Oyedokun, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Rull, S., Savary, G., Rivers, M., Rojas, A., Rojokan, I., Rolle, M., Ruiz, S., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sison, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swalek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umanak, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Washington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Unpublished
2 (bases 1 to 205825)

Direct Submission
Submitted (09-AUG-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jan 2, 2002 this sequence version replaced gi:15135816.

----- Genome Center -----

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: RP11-479J2
Center clone name: RP11-479J2
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329first call to findPhrapList
Consensus quality: 198598 bases at least Q40
Consensus quality: 208850 bases at least Q30
Consensus quality: 214667 bases at least Q20
Estimated insert size: 207320; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 3.5x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
26477: contig of 26477 bp in length
26478
26578: gap of unknown length
49296: contig of 22719 bp in length
49297
49396: gap of unknown length
68898: contig of 19502 bp in length
68899
68998: gap of unknown length
86550: contig of 17552 bp in length
86551
86650: gap of unknown length
100899: contig of 14249 bp in length
100900
100999: gap of unknown length
112284: contig of 11285 bp in length
112285
112384: gap of unknown length
121682: contig of 9298 bp in length
121683
121782: gap of unknown length
135703: contig of 13921 bp in length
135704
135803: gap of unknown length
135804
145468: contig of 9665 bp in length
145469
145468: gap of unknown length
154983: contig of 9415 bp in length
154984
155083: gap of unknown length

155084 164549: contig of 9466 bp in length
164550 164649: gap of unknown length
164650 171823: contig of 7174 bp in length
171824 171923: gap of unknown length
171924 180207: contig of 8284 bp in length
180208 180307: gap of unknown length
180308 187008: contig of 6701 bp in length
187009 187108: gap of unknown length
187109 191344: contig of 4236 bp in length
191345 191444: gap of unknown length
191445 195679: contig of 4235 bp in length
195680 195779: gap of unknown length
195780 199345: contig of 3566 bp in length
199346 199445: gap of unknown length
203512 203512: contig of 4067 bp in length
203513 203612: gap of unknown length
203613 205825: contig of 2213 bp in length.

FEATURES
source 1..205825
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q"
/clone="RP11-479J2"

BASE COUNT 63155 a 40250 c 40422 g 60165 t 1833 others
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AC092942 ..
Align seg 1/1 to: AC092942 from: 1 to: 205825

8 LeuThrIleSerAlaLeuThrAla 16
|||||
1010 CTCACATCTCTCTCTCTCGACGCA 1036

seq_name: gb_hgtc:AC099581

seq_documentation_block:
LOCUS AC099581 209518 bp DNA linear HTG 16-NOV-2001
DEFINITION Mus musculus chromosome 4 clone RP23-184J13 map 4, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
ACCESSION AC099581
VERSION AC099581.1 GI:16946030
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTROP.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 209518)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus chromosome 4, clone RP23-184J13
Unpublished
2 (bases 1 to 209518)
Anderson, S., Barra, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, V., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Grinde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kanat, A., Karatas, A., Kells, C., Lacroque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPherson, R., Meldrum, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,

TITLE JOURNAL COMMENT

Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schuppach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Topham, K., Travers, M., Talamas, J., Testaye, S., Theodore, J., Tripathi, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (16-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WTHR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project information
Center project name: L16751
Center clone name: 184.J.13

----- Summary Statistics -----
Sequencing vector: plasmid; R/A; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 206496 bases at least Q40
Consensus quality: 207387 bases at least Q40
Consensus quality: 207727 bases at least Q20
Insert size: 214000; agarose-fp
Insert size: 208318; sum-of-contigs
Quality coverage: 8.8 in Q20 bases; agarose-fp
Quality coverage: 9.1 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 12457: contig of 12457 bp in length
12458 12557: gap of 100 bp
12558 13208: contig of 651 bp in length
13209 13308: gap of 100 bp
13309 14078: contig of 770 bp in length
14079 14178: gap of 100 bp
14179 16566: contig of 2388 bp in length
16567 16666: gap of 100 bp
16667 20489: contig of 3823 bp in length
20490 20589: gap of 100 bp
20590 25395: contig of 4806 bp in length
25396 25495: gap of 100 bp
25496 32724: contig of 7229 bp in length
32725 32824: gap of 100 bp
32825 42342: contig of 9518 bp in length
42343 42442: gap of 100 bp
42443 49596: contig of 7154 bp in length
49597 49696: gap of 100 bp
49697 67428: contig of 17733 bp in length
67429 67528: gap of 100 bp
67529 102738: contig of 35209 bp in length
102739 102838: gap of 100 bp
102839 172277: contig of 69439 bp in length
172278 172377: gap of 100 bp
172378 209518: contig of 37141 bp in length.
Location/Qualifiers
1. 209518

```

FEATURES
source
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/map="4"

```

misc_feature      /clone="Rp23-184J13"
                  /clone_id="RC1-23 Female Mouse BAC"
                  1. 12457
                  /note="assembly-fragment"
                  clone_end:SP6
                  vector_side:left"
misc_feature      12558..13208
                  /note="assembly-fragment"
misc_feature      13309..14078
                  /note="assembly-fragment"
misc_feature      14179..16566
                  /note="assembly-fragment"
misc_feature      16667..20489
                  /note="assembly-fragment"
misc_feature      20590..25395
                  /note="assembly-fragment"
misc_feature      25496..32724
                  /note="assembly-fragment"
misc_feature      32825..42342
                  /note="assembly-fragment"
misc_feature      42443..49596
                  /note="assembly-fragment"
misc_feature      49697..67428
                  /note="assembly-fragment"
misc_feature      67530..102738
                  /note="assembly-fragment"
misc_feature      102839..172277
                  /note="assembly-fragment"
misc_feature      172378..209518
                  /note="assembly-fragment"
                  clone_end:T7
                  vector_side:right"
BASE COUNT      55725 a 44920 c 47112 g 60522 t 1239 others
ORIGIN

```

```

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

```

alignment_block:
US-10-048-197-2 x AC099581 ..
Align seg 1/1 to: AC099581 from: 1 to: 209518

```

```

12 AlaleuLeuThraLeuDeuValThr 20
|||||
5884 GCCCTCCTCACAGCAGCTAGTTTACA 5910

```

```

seq_name: gb_ba:AP003590

```

```

seq_documentation_block:

```

```

LOCUS      AP003590      333500 bp      DNA      linear      BCT 28-NOV-2001
DEFINITION Nostoc sp. PCC 7120 DNA, complete genome, section 10/19.
ACCESSION  AP003590 BA000019
VERSION    AP003590.1 GI:17131676
KEYWORDS

```

```

SOURCE     Nostoc sp. PCC 7120 DNA.
           Nostoc sp. PCC 7120

```

```

ORGANISM   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
1 (sites)
AUTHORS    Kaneko, T., Nakamura, Y., Molk, C. P., Kuritz, T., Sasamoto, S.,
           Watanabe, A., Iriyuchi, M., Ishikawa, A., Kawashima, K., Kimura, T.,
           Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A.,
           Nakazaki, N., Shimo, S., Sugimoto, M., Takazawa, M., Yamada, M.,
           Yasuda, M. and Yabata, S.

```

```

TITLE      Complete genomic sequence of the filamentous nitrogen-fixing
           cyanobacterium Anabaena sp. strain PCC 7120
JOURNAL    DNA Res. 8 (5), 205-213 (2001).
MEDLINE    21595285
PUBMED     11759840

```

```

REFERENCE  2 (bases 1 to 333500)

```

AUTHORS

TITLE

JOURNAL

Kaneko, T.
 Direct Submission
 Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research
 Institute, The First Laboratory for Plant Gene Research, Yana
 1532-3, Kisarazu, Chiba 292-0812, Japan
 (E-mail: kaneko@kazusa.or.jp,
 URL: http://www.kazusa.or.jp/cyanobase/
 Tel: 81-438-52-3935 (ex. 338), Fax: 81-438-52-3934)

FEATURES

source

1. .333500
 Location/Qualifiers

/organism="Nostoc sp. PCC 7120"

/db_xref="taxon:103690"

/note="synonym: Anabaena sp. PCC7120"

/complement(57. .926)

/gene="a112584"

/complement(57. .926)

/gene="a112584"

/note="ORF_ID:a112584"

/codon_start=1

/transl_table=11

/product="ATP-binding protein of ferrichrome ABC
 transporter"

/protein_id="BAB74283.1"

/db_xref="GI:1713167"

/translation="MSVKPARLTLNDSHSTVNTOKIANKKGLSTOSISLAYDGPVPIIR
 DNLAIPGKISALVANGCGKSTLRGLARLKPYGTVHLDGOSIFSLATSEVAKO
 LGILPGVPAREGLTYKDYVGGSRPYQNMLOQMSKDEIYQOALEITNLSLADRA
 LDTLSGGQKORAMIAALADDTIDLLDEPTTFDLAHQLEVDLLYEINQHERIV
 MVLHDNQACRADYLVVYKGRIFTAGEPKDVTEMEQVEVFGQCRIVSDPVGTP
 MCVPIGRKGEKO"

/complement(964. .2004)

/gene="a112585"

/complement(964. .2004)

/gene="a112585"

/note="ORF_ID:a112585"

/codon_start=1

/transl_table=11

/product="Iron(III) dicitrate transport system permease
 protein"

/protein_id="BAB74284.1"

/db_xref="GI:17131678"

/translation="MKVDMVLRSETMSLRIDRRVPIILCLVVAIVAVMNNLGRGE
 YPIAPDIKTIVGIDTGNPDHAFVYINLRPLTVACMYGVALAVGTFIOGTRNP
 LADPGIIGINAGASLAAMVYIVFSPASPIYTLPLSAPAGALLMAGLIYSIAMNNGSSP
 VLPILMGVGLSATAAGAFSLMTIFGDIYSISDLVWLACSVYGRTEBOYFSFPIMLIV
 FVPMLATLRNHLNTLVGDVAGLSTRVEMQGLLVGVVALAGAVATAGMIGEVG
 LIAPIHLGRQLVGTNHOGLPLPTSAIAGMLVYVADFLGRTLFAPRIEIPCGVYTAAGAP
 YELYLIRNRKK"

/complement(2001. .3032)

/gene="a112586"

/complement(2001. .3032)

/gene="a112586"

/note="ORF_ID:a112586"

/codon_start=1

/transl_table=11

/product="Iron(III) dicitrate transport system permease
 protein"

/protein_id="BAB74285.1"

/db_xref="GI:17131679"

/translation="MTKATATASPRNNPKISPLVGLIIGILILILCLVSVYTLGAEE
 IPLNLTSEITFDGSEYHLYIQTVLRSLVALLGVSSLAVALMOGLTRPLADP
 GILGIESGALAVTTITFVSGSSGLITTVAFGLGVQVPMILVYPIGSGKGGATPLN
 LTYAGALALISLTLTALIVSQRTLEERFLAGSLAGRDNIILSLAPYPMIGLY
 VAFALGRQITMSLGEDMAKGLDQOTAWKITTIAISVVLGSSVSLAGTIGTGLV
 PHMVRPFIKADYKWLIPYSVAVGATILLVADVAARVLAKFQELPVGVMALVGAPEFY
 YLAKSVYKK"

/complement(2001. .3032)

/gene="a112587"

/complement(2001. .3032)

/gene="a112587"

/note="ORF_ID:a112587"

/codon_start=1

/transl_table=11

/product="Iron(III) dicitrate transport system permease
 protein"

/protein_id="BAB74286.1"

/db_xref="GI:17131679"

/translation="MTKATATASPRNNPKISPLVGLIIGILILILCLVSVYTLGAEE
 IPLNLTSEITFDGSEYHLYIQTVLRSLVALLGVSSLAVALMOGLTRPLADP
 GILGIESGALAVTTITFVSGSSGLITTVAFGLGVQVPMILVYPIGSGKGGATPLN
 LTYAGALALISLTLTALIVSQRTLEERFLAGSLAGRDNIILSLAPYPMIGLY
 VAFALGRQITMSLGEDMAKGLDQOTAWKITTIAISVVLGSSVSLAGTIGTGLV
 PHMVRPFIKADYKWLIPYSVAVGATILLVADVAARVLAKFQELPVGVMALVGAPEFY
 YLAKSVYKK"

/complement(2001. .3032)

/gene="a112588"

/complement(2001. .3032)

/gene="a112588"

/note="ORF_ID:a112588"

/codon_start=1

/transl_table=11

/product="Iron(III) dicitrate transport system permease
 protein"

/protein_id="BAB74287.1"

/db_xref="GI:17131679"

/translation="MTKATATASPRNNPKISPLVGLIIGILILILCLVSVYTLGAEE
 IPLNLTSEITFDGSEYHLYIQTVLRSLVALLGVSSLAVALMOGLTRPLADP
 GILGIESGALAVTTITFVSGSSGLITTVAFGLGVQVPMILVYPIGSGKGGATPLN
 LTYAGALALISLTLTALIVSQRTLEERFLAGSLAGRDNIILSLAPYPMIGLY
 VAFALGRQITMSLGEDMAKGLDQOTAWKITTIAISVVLGSSVSLAGTIGTGLV
 PHMVRPFIKADYKWLIPYSVAVGATILLVADVAARVLAKFQELPVGVMALVGAPEFY
 YLAKSVYKK"

/complement(2001. .3032)

/gene="a112589"

/complement(2001. .3032)

/gene="a112589"

/note="ORF_ID:a112589"

/codon_start=1

/transl_table=11

/product="transcriptional regulator"

/protein_id="BAB74286.1"

/db_xref="GI:17131680"

/translation="MYLIQESADCYQSLTENNTPQSNWEHIDPOGMSHCYROM
 FHLRPGISLTDHYQLNNDLIVKISAPNCLMWLEFSIFGNNNNDVYSGQNFVYR
 RELGKAGYKMQAQRILKDFDIEYSDQNLALHOMDYIPVSROMPEKNDPEYR
 NFGITTHEMOTVLRQILNCPYGLQOLYEGKILLELALRLYLENPFNKSTLK
 PHQDSIVHARQDILINNPPSLINLAROVGLNDCOTLKKGFEIGTGYFGYLHBYR
 MEAKELLIDKKNVQVAOIVGYEARTSTIRAFKKFGVSPAYVQKNSA"

/complement(4640. .7195)

/gene="a1r2588"

/complement(4640. .7195)

/gene="a1r2588"

/note="ORF_ID:a1r2588"

/codon_start=1

/transl_table=11

/product="ferrichrome-iron receptor"

/protein_id="BAB74287.1"

/db_xref="GI:17131681"

/translation="MRGEMLIKPALHRLCLATVSVLSISNOTVMAEVMNTSKTPOL
 SKQVPILAQPTTEILEVYGVKANPTEKCEVYLQTTSEKQLQVNRSGNAYIADIP
 MAOLRLNGDTFTFTSPKPEYEGISEIIVSLMDANTLRVTVGAGLPEVLEFSDGL
 ILGVLSPASQQTATPTPEEPQAAITTDQPIELVVGQDSKRVTAATAATKDTPLR
 DLPQSIQVIRQVIEDQRAQTLNEALRNVSQVPSASGRTDRFVIRGFDPNSVIR
 DGRKENTSVYREANIERIEVLKGPASVLEGOPEGVINVIKOPQRPFLMGLEA
 GSYGFRPRPYDENSPLNDSKTLRLRVNAVEISSEPRDDKERSRFFIAPVLAWEIGD
 RSLVVDLETLRSPDRGLVALGQRAVEIPNRNLGEPDSKEVEDLRICTREHIG
 FDNMKLRSAFRIVSTQSSVTEPGSLDQATLSDRVDPTDEYAFOTDLIG
 KFKTGAIEHVLGFDFNROTINFDNRDSAPAAIDIFNVDYLTARPNIESDPRLIF
 AADNIGIYIODOIKLAEMLKILGIRYDFNQSFISIVDGKFFDLSQAFSPRIGIV
 YQITPPLSYASYSRSPQNFQGIORNSRIEPRRGQYEVGERTLDGLSLAGIV
 QITKTVNAPDPAIDINESTIPGCVSRSGRIEFDIAGLAKGMNIIASVATDAKITEDN
 TDNEGRLNRVNPNSASIVTTELOSGALDGLMGVGLFFVGGRQDLSNFPVPGYT
 RIDAALFRRDNNNIGLINFKNIPDVNYIESTISRTSIDAGIPTVIGLSVKK"

/complement(7241. .8194)

/gene="a1r2589"

/complement(7241. .8194)

/gene="a1r2589"

/note="ORF_ID:a1r2589"

/codon_start=1

/transl_table=11

/product="ATP-binding protein of ferrichrome ABC
 transporter"

/protein_id="BAB74283.1"

/db_xref="GI:1713167"

/translation="MSVKPARLTLNDSHSTVNTOKIANKKGLSTOSISLAYDGPVPIIR
 DNLAIPGKISALVANGCGKSTLRGLARLKPYGTVHLDGOSIFSLATSEVAKO
 LGILPGVPAREGLTYKDYVGGSRPYQNMLOQMSKDEIYQOALEITNLSLADRA
 LDTLSGGQKORAMIAALADDTIDLLDEPTTFDLAHQLEVDLLYEINQHERIV
 MVLHDNQACRADYLVVYKGRIFTAGEPKDVTEMEQVEVFGQCRIVSDPVGTP
 MCVPIGRKGEKO"

/complement(964. .2004)

/gene="a112585"

/complement(964. .2004)

/gene="a112585"

/note="ORF_ID:a112585"

/codon_start=1

/transl_table=11

/product="Iron(III) dicitrate transport system permease
 protein"

/protein_id="BAB74284.1"

/db_xref="GI:17131678"

/translation="MKVDMVLRSETMSLRIDRRVPIILCLVVAIVAVMNNLGRGE
 YPIAPDIKTIVGIDTGNPDHAFVYINLRPLTVACMYGVALAVGTFIOGTRNP
 LADPGIIGINAGASLAAMVYIVFSPASPIYTLPLSAPAGALLMAGLIYSIAMNNGSSP
 VLPILMGVGLSATAAGAFSLMTIFGDIYSISDLVWLACSVYGRTEBOYFSFPIMLIV
 FVPMLATLRNHLNTLVGDVAGLSTRVEMQGLLVGVVALAGAVATAGMIGEVG
 LIAPIHLGRQLVGTNHOGLPLPTSAIAGMLVYVADFLGRTLFAPRIEIPCGVYTAAGAP
 YELYLIRNRKK"

/complement(2001. .3032)

/gene="a112586"

/complement(2001. .3032)

/gene="a112586"

/note="ORF_ID:a112586"

/codon_start=1

/transl_table=11

/product="Iron(III) dicitrate transport system permease
 protein"

/protein_id="BAB74285.1"

/db_xref="GI:17131679"

/translation="MTKATATASPRNNPKISPLVGLIIGILILILCLVSVYTLGAEE
 IPLNLTSEITFDGSEYHLYIQTVLRSLVALLGVSSLAVALMOGLTRPLADP
 GILGIESGALAVTTITFVSGSSGLITTVAFGLGVQVPMILVYPIGSGKGGATPLN
 LTYAGALALISLTLTALIVSQRTLEERFLAGSLAGRDNIILSLAPYPMIGLY
 VAFALGRQITMSLGEDMAKGLDQOTAWKITTIAISVVLGSSVSLAGTIGTGLV
 PHMVRPFIKADYKWLIPYSVAVGATILLVADVAARVLAKFQELPVGVMALVGAPEFY
 YLAKSVYKK"

/complement(2001. .3032)

/gene="a112587"

/complement(2001. .3032)

/gene="a112587"

/note="ORF_ID:a112587"

/codon_start=1

/transl_table=11

/product="Iron(III) dicitrate transport system permease
 protein"

/protein_id="BAB74286.1"

/db_xref="GI:17131679"

/translation="MTKATATASPRNNPKISPLVGLIIGILILILCLVSVYTLGAEE
 IPLNLTSEITFDGSEYHLYIQTVLRSLVALLGVSSLAVALMOGLTRPLADP
 GILGIESGALAVTTITFVSGSSGLITTVAFGLGVQVPMILVYPIGSGKGGATPLN
 LTYAGALALISLTLTALIVSQRTLEERFLAGSLAGRDNIILSLAPYPMIGLY
 VAFALGRQITMSLGEDMAKGLDQOTAWKITTIAISVVLGSSVSLAGTIGTGLV
 PHMVRPFIKADYKWLIPYSVAVGATILLVADVAARVLAKFQELPVGVMALVGAPEFY
 YLAKSVYKK"

/complement(2001. .3032)

/gene="a112588"

/complement(2001. .3032)

/gene="a112588"

/note="ORF_ID:a112588"

/codon_start=1

/transl_table=11

/product="Iron(III) dicitrate transport system permease
 protein"

/protein_id="BAB74287.1"

/db_xref="GI:17131679"

/translation="MTKATATASPRNNPKISPLVGLIIGILILILCLVSVYTLGAEE
 IPLNLTSEITFDGSEYHLYIQTVLRSLVALLGVSSLAVALMOGLTRPLADP
 GILGIESGALAVTTITFVSGSSGLITTVAFGLGVQVPMILVYPIGSGKGGATPLN
 LTYAGALALISLTLTALIVSQRTLEERFLAGSLAGRDNIILSLAPYPMIGLY
 VAFALGRQITMSLGEDMAKGLDQOTAWKITTIAISVVLGSSVSLAGTIGTGLV
 PHMVRPFIKADYKWLIPYSVAVGATILLVADVAARVLAKFQELPVGVMALVGAPEFY
 YLAKSVYKK"

/complement(2001. .3032)

/gene="a112589"

/complement(2001. .3032)

/gene="a112589"

/note="ORF_ID:a112589"

/codon_start=1

/transl_table=11


```

/product="transcriptional regulator"
/protein_id="BAB74290.1"
/db_xref="GI:17131684"
/translation="MYLILQESADCYQPIPTENNNTPOSTIMEAHITHDPQMSHGKROM
FHLRPGISLITDCYQINDDLVVEKAPSPALTWELSPNFGNIGEGYPCGNFIDYV
KGLPEANHTEMOAEKVKFIDHLESELOPLFAHOMLPLSPFRMEKNDDEPR
NFGITHEMOIVLROILNCPYQGLTKOGLBEKILFELALREYQENPTINTDNL
KNHVDGIYHARDILIKNLNPPSLIDLARLVGLNDCTLKGFQDEFTAFGLIHEX
RMEQAKELLEEMVTVQAVGVYEARTSPFRARKNFVSPTAVIQRNSA"
10585..13134
/gene="alr2592"
/gene="alr2592"
10585..13134
/gene="alr2592"
/contig="alr2592"
/codon_start=1
/translation="11
/product="ferrichrome-iron receptor"
/protein_id="BAB74291.1"
/db_xref="GI:17131685"
/translation="MLIKAGLNLTYLTIATVSIISNQVRAEVNTGNGOTPOLSKDV
SILAQPTTEIVQVYANPTKEGVYILOTGSEKLOLVNLSQKAYITDIPNKL
RLPNDSTFTSPKPEVGEISLITVINDANTIRVTGCGTLPVELFDSDELIGV
VSATPQOTAEQPAQOTENOEPIELVYGEDITYSNASTATRTDPLDIPQS
IOVLPQVLRKQOYTRILDARVNSGVTPOKGIQDAIDFTTIGFNSNRRLRNGFNR
AARGSTFTAPSSVEVEVVKPAPSLVQLEGGLVNTVTKPLSDPYAAEFYAG
SYSPSPSIDISGLTTDKLILRLNLSYNGFSFIDFVNGENLSIYPVNYQISDR
ALTEPERSKNSQSFNDGLPDIISFELISRLNNDPDDLTNTSLNFTLEHGFND
NIRLRTVFGASFEGTILAYRLFNYNENETNEVRAFNSPDYENDYSMODVAKRNT
GAVRHELGVFELAFNDVNAEDSEDSFINVFNPDPIPLATYSFGSKSNNTIG
LYLQDQITSLPNKILIGRRYDFARAKSDSNPQIFGEPSTDEYNNAFSRSVIG
YQPEISLVSASRSLPNNAKTGEITTEPRKQLEGVADGRISTLAAEI
TKNIIISDRNDPEFSGEVSRGIEDISELLPGMNVASTPINSVITODNML
PIGDTLVNAPRRKASLMTTETIQDQIGFGLGVYVGDROAELNNMLVPSYVRA
DASIFYRDMRIGLNKKNLSDTRYESSONSLIYGPPLVLTGVSYQF"

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AP003590/rev ..

Align seg 1/1 to reverse of: AP003590 from: 1 to: 333500

85 IlleusSerleullelleSerphelen 93
|||||
276344 ATATTATCTTTAATTATCTCATCTCTA 276318

seq_name: gb_pat:AX120599

seq_documentation_block:
LOCUS AX120599 261 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 515 from Patent EPI108790.
ACCESSION AX120599
VERSION AX120599.1 GI:14037314
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 261)
Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
Novel polynucleotides
Patent: EP 1108790-A 515 20-JUN-2001;
JOURNAL KYOKA HAKO KOGYO CO., LTD. (JP)
FEATURES
Location/Qualifiers
1..261
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
BASE COUNT 52 a 77 c 54 g 78 t


```

```

ORIGIN
alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AX120599 ..

Align seg 1/1 to: AX120599 from: 1 to: 261

11 SerAlaLeuLeuThrAlaLeuLeu 18
|||||
95 TCCGGCTTGCTCACAGCCCTTCTT 118

seq_name: gb_pl:AF314550

seq_documentation_block:
LOCUS AF314550 293 bp mRNA linear PLN 02-NOV-2001
DEFINITION Glycine max clone NTN5c NBS-type putative resistance protein mRNA,
partial cds.
ACCESSION AF314550
VERSION AF314550.1 GI:16588591
KEYWORDS
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE 1 (bases 1 to 293)
Jeong,S.C., Hayes,A.J., Blyashov,R.M. and Saghai Maroof,M.A.
Diversity and evolution of a non-TIR-NBS sequence family that
clusters to a chromosomal 'hotspot' for disease resistance genes in
soybean
Unpublished
2 (bases 1 to 293)
Jeong,S.C., Hayes,A.J., Blyashov,R.M. and Saghai Maroof,M.A.
Direct Submission
Submitted (18-OCT-2000) Crop and Soil Environmental Sciences,
Virginia Tech, 350 Smyth Hall, Blacksburg, VA 24061-0404, USA
FEATURES
Location/Qualifiers
1..293
/organism="Glycine max"
/db_xref="taxon:3847"
<1..>293
/codon_start=2
/product="NBS-type putative resistance protein"
/protein_id="AA126851.1"
/db_xref="GI:16588592"
/translation="MGVGKTTIAQHVYNDPRIEGKFDIKAMCVSDPDVLTVTTRAI
LEAVIDSTNSRGLMEVHRRLKENIGKRFLLVLDVVDVNNERKREAVLTPIN"
BASE COUNT 89 a 41 c 79 g 84 t
ORIGIN

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AF314550 ..

Align seg 1/1 to: AF314550 from: 1 to: 293

50 GlnGluIleSerSerArgpHeGly 57
|||||
214 CAGAGATTCTCTCTGTTTGGA 237


```



```

seq_name: gb_pl:AF222879
seq_documentation_block:
LOCUS AF222879 312 bp DNA linear PLN 15-MAR-2000
DEFINITION Glycine max clone R14 disease resistance-like protein gene, partial
cds.
ACCESSION AF222879
VERSION AF222879.1 GI:7243628
KEYWORDS
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE
1 (bases 1 to 312)
Hayes, A.J. and Saghai Maroof, M.A.
Targeted resistance gene mapping in soybean using modified AFLPs
JOURNAL Theor. Appl. Genet. (2000) In press
AUTHORS Hayes, A.J. and Saghai Maroof, M.A.
REFERENCE
2 (bases 1 to 312)
Direct Submission
Submitted (10-JAN-2000) Crop and Soil Environmental Sciences,
Virginia Tech, Blacksburg, VA 24061-0404, USA
FEATURES
source
1..312
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="R14"
<1..>312
/codon_start=1
/product="disease resistance-like protein"
/protein_id="AAFA3387.1"
/db_xref="GI:7243629"
/translation="GMGCGKTTTAAQHYNDPRIGKEDIKAMVCSDPDVLTVTTRA
ILEAVIDSTDSNRGLEWHRRLEKILGRFLVDDVWNEKREMEAVQPTLYGAR
GS"
BASE COUNT 90 a 46 c 89 g 87 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AF222879 ..
Align seg 1/1 to: AF222879 from: 1 to: 312
50 GlnGluIleSerSerArgPheGly 57
|||||
216 CAAGAGATTCTCTCTCGTTTGA 239

seq_name: gb_in:TIU42193
seq_documentation_block:
LOCUS TIU42193 407 bp DNA linear INV 13-MAR-1996
DEFINITION Taeniotrichips inconsequens cytochrome oxidase I (COI) gene,
mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION U42193
VERSION U42193.1 GI:1223995
KEYWORDS
SOURCE Taeniotrichips inconsequens.
ORGANISM Taeniotrichips inconsequens
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Paraneoptera; Thysanoptera; Terebrantia;
Thripidae; Thripidae; Taeniotrichips.
REFERENCE
1 (bases 1 to 407)
Crespi, B.J., Carmean, D.A., Vawter, L. and von Dohlen, C.
Molecular Phylogenetics of Thysanoptera
JOURNAL Systematic Entomology (1996) In press

```

```

REFERENCE
2 (bases 1 to 407)
Carmean, D.A., Crespi, B.J., Vawter, L. and von Dohlen, C.
Direct Submission
Submitted (06-DEC-1995) David A. Carmean, Biological Sciences,
Simon Fraser University, Burnaby, BC V5A1S6, Canada
FEATURES
source
1..407
/organism="Taeniotrichips inconsequens"
/organeller="mitochondrion"
/db_xref="taxon:45055"
1..407
/gene="CO1"
<1..>407
/gene="CO1"
/codon_start=3
/transl_table=5
/product="cytochrome oxidase I"
/protein_id="AAAG92103.1"
/db_xref="GI:1223996"
/translation="LPPSITILLIMGLSKREGAGTGWTVVPPSTFYHSGISVDLTIFSL
HLGISSILGALNFTTITMNLKIKNLSKSTYLFVWSVILTAIILLSLPVLGATTM
LTDRNLNLSFDPGCGDPVLYOHLFWFFGHP"
BASE COUNT 126 a 75 c 52 g 154 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x TIU42193/rev ..
Align seg 1/1 to reverse of: TIU42193 from: 1 to: 407
64 PheMetIleValIleIleYsPhe 71
|||||
193 TTTATGATTGTAGTAAATATTT 170

seq_name: gb_pl:AF060192
seq_documentation_block:
LOCUS AF060192 519 bp DNA linear PLN 23-FEB-2001
DEFINITION Glycine max putative resistance protein KNBS4 gene, partial cds.
ACCESSION AF060192 AF325687
VERSION AF060192.2 GI:13111696
KEYWORDS
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE
1 (bases 1 to 519)
He, C. and Chen, S.Y.
Resistance gene analogs from soybean
JOURNAL Unpublished
2 (bases 1 to 519)
Dong, W. and Chen, S.Y.
Direct Submission
Submitted (16-APR-1998) 803 Lab, Institute of Genetics, CAS,
Beijing 100101, P.R. China
3 (bases 1 to 519)
He, C. and Chen, S.Y.
Direct Submission
Submitted (23-FEB-2001) Plant Biotechnology Laboratory, Institute
of Genetics, CAS, Datun Road, Beijing 100101, P.R. China
REMARK
Sequence update by submitter.
On Feb 23, 2001 this sequence version replaced gi:3091215.
COMMENT
Location/Qualifiers
1..519
/organism="Glycine max"

```

```

mRNA      /db_xref="taxon:3847"
<1..>519
/Product="putative resistance protein KMS4"
CDS       <1..>519
/codon_start=1
/product="putative resistance protein KMS4"
/protein_id="AAC15224.2"
/db_xref="GI:13111697"
/translation="SGVGKTTLAOHVYSPDRIGKRVIKAWCVSDPDVLTVTNRI
LEAVIDSTDSNRSGLSEWVHRRIKRIKRLIVLDDYVKNRKEKREKLVLPITVGAR
SRIVLTFTTSTKVASVTSRKNKRLHLEQJEDHCWKVFAKFAFODDNPRLNKLKIGIK
IVERCKGIPPLAG"
BASE COUNT 155 a 79 c 139 g 146 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AF060192 ..
Align seg 1/1 to: AF060192 from: 1 to: 519
50 GINGLIIIESerSeraPhegly 57
|||||
213 CAAGAGATTCTTCCTTCGTTTGA 236

seq_name: gb_pl:AB027824

seq_documentation_block:
LOCUS AB027824 521 bp DNA linear PLN 21-JUN-2001
DEFINITION Schizosaccharomyces pombe gene for Protein kinase C-like 2, partial
cds, clone: S644.
ACCESSION AB027824
VERSION AB027824.1 GI:6478831
KEYWORDS Schizosaccharomyces pombe (strain:968 h90) DNA, clone:S644.
SOURCE Schizosaccharomycetes pombe
ORGANISM Schizosaccharomycetes pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
1 (sites)
REFERENCE Ding,D.O., Tomita,Y., Yamamoto,A., Chikashige,Y., Haraguchi,T. and
AUTHORS Hiraoka,Y.
TITLE Large-scale screening of intracellular protein localization in
living fission yeast cells by the use of a GFP-fusion genomic DNA
library
JOURNAL Genes Cells 5 (3), 169-190 (2000)
2 (bases 1 to 521)
Ding,D., Tomita,Y. and Hiraoka,Y.
Direct Submission
Submitted (24-MAY-1999) Da-Oiao Ding, Communications Research
Laboratory, Kansai Advanced Research Center: 588-2, Iwoka,
Iwoka-cho, Nishiku, Kobe, Hyogo 651-2401, Japan
(E-mail:dinge@rcl.go.jp, Tel:+81-78-969-2240, Fax:+81-78-969-2249)
An S. pombe gene library in which genomic DNA fragments are fused
to the 5'-end of the GFP-S65T gene was constructed. S. pombe strain
was transformed with the library DNA amplified in E.coli. Plasmids
of those transformants that showed interesting GFP localization
detailed information and images can be seen in our web site
(http://www-karc.rcl.go.jp/bio/cellmagic/). Note: Only major
localization is described for each clone in most cases. Because it
is a random fusion of DNA fragment with the GFP gene, the
subcellular localization of the fusion protein may differ from that
of the full-length gene product.
Location/Qualifiers
1..521
/organism="Schizosaccharomyces pombe"
FEATURES
Source

```

```

/strain="968 h90"
/db_xref="taxon:4896"
/clone="S644"
/note="subcellular localization of GFP fusion; Cell
periphery and septum"
1..521
/gene="pck2+"
<1..>521
/gene="pck2+"
/codon_start=1
/product="Protein kinase C-like 2"
/protein_id="BAA87128.1"
/db_xref="GI:6478832"
/translation="TETRIYFVMDVFGSGDMLHIOQEFSSRRRQFYAAEVCALAKY
FHDNGIIRYRDLKNDLILSPDGHYKADYGLCKRDMHNDITAFCGTPEFMAPEILL
EQGYTRSYDMMARFGLYITOMLLGQSPRGEDEERIFDAILSDPELYPIHMPRDSVIL
QQLTRDPKRLGS"
BASE COUNT 146 a 93 c 111 g 171 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AB027824 ..
Align seg 1/1 to: AB027824 from: 1 to: 521
85 IIELeuSerLeuIIeIleSerPhe 92
|||||
161 ATCTTAAGCTTGATATATACCTTT 184

seq_name: gb_pat:AX317918

seq_documentation_block:
LOCUS AX317918 562 bp DNA linear PAT 14-DEC-2001
DEFINITION Sequence 19 from Patent WO0190149.
ACCESSION AX317918
VERSION AX317918.1 GI:17900772
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
REFERENCE Vogeli,G.
AUTHORS G protein-coupled receptors
TITLE Patent: WO 0190149-A 19 29-NOV-2001;
JOURNAL PHARMACIA & UPHOHN COMPANY (US)
FEATURES
Source
1..562
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 176 a 118 c 140 g 128 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AX317918/rev ..
Align seg 1/1 to reverse of: AX317918 from: 1 to: 562
10 IIESeRAlaLeuLeuThraLaLeu 17
|||||
66 ATTCTGCGCTTCTCAGTCCTTG 43

```

```

seq_name: gb_pl:AY063021
seq_documentation_block:
LOCUS      AY063021                847 bp    mRNA    linear    PLN 26-NOV-2001
DEFINITION Arabidopsis thaliana putative myo-inositol monophosphatase
            (F13E7_19/AT3g02870) mRNA, complete cds.
ACCESSION  AY063021
VERSION     AY063021.1  GI:17104612
KEYWORDS   F11_CDNA.
SOURCE      thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
            1 (bases 1 to 847)
REFERENCE  1
AUTHORS    Yamada,K., Banno,F., Banno,F., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamamura,Y., Yu,G., Yu,S., Bowser,L., Carninci,P., Chen,H.,
            Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Kim,C., Koeseema,E., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tracy,S.E.,
            Shinozaki,K., Davis,R.W., Becker,J.R. and Theologis,A.
            Full Length cDNA of gene F13E7_19/AT3g02870 (GI:6728975)
            Unpublished
            2 (bases 1 to 847)
REFERENCE  2
AUTHORS    Yamada,K., Banno,F., Banno,F., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamamura,Y., Yu,G., Yu,S., Bowser,L., Carninci,P., Chen,H.,
            Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Kim,C., Koeseema,E., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tracy,S.E.,
            Shinozaki,K., Davis,R.W., Becker,J.R. and Theologis,A.
            Direct Submission
            Submitted (15-NOV-2001) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     The RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RFL CDNA (RFL CDNA: 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.
            The Salk, Stanford, PGEC (SSP) Consortium members constructed and
            sequenced the pUNT (ORF) clones using the RFL CDNA: Yamada,K.,
            Banno,F., Banno,F., Chang,E., Dale,J.M., Goldsmith,A.D., Lee,J.M.,
            Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamamura,Y., Yu,G., Yu,S., Bowser,L., Carninci,P., Chen,H.,
            Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Kim,C., Koeseema,E., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A.,
            Tracy,S.E., Davis,R.W., Becker,J.R. and Theologis,A.
FEATURES
source
1..847
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="3"
/clone="U09692"
/notes="This clone is in pUNT 51.
ecotype: Columbia"
1..847
/gene="F13E7_19/AT3g02870"
1..816
/gene="F13E7_19/AT3g02870"
/codon_start=1
/evidence="experimental
/product="putative myo-inositol monophosphatase"
/protein_id="AA134195.1"
/db_xref="GI:17104613"

```

```

/translation="MADNDSLPDPIAAIDAKKAGQIIRKGEYETKHVEHKGOVDLY
TFEDKGCCEIYFNHILKOLEPNNKFTIGEETTAFCWTETLDEPTWIVDPLDGTNTNPHG
FFPCVCSISLITIGKVPVGVYVNPINPEELFTYVGKGAFLNKRIRKVSQSLRLALL
VTEAGTRKQATLDPTNRINSLTRKVRSLRNSGSCALDLCGACGRVDIETELGFGG
PMDIAGIYIVEAGGLIFDPGKIDITSORIAASMSLKELFALRLTGA"
3'UTR
BASE COUNT      231 a      164 c      215 g      237 t
ORIGIN

```

```

alignment_scores:
Quality:      8.00      Length:      8
Ratio:      1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

```

```

alignment_block:
US-10-048-197-2 x AY063021 ..
Align seg 1/1 to: AY063021 from: 1 to: 847

```

```

13 LeuleuthralaleuleuValThr 20
|||||
463 CTTTAAACCGCTTGTCTGTGACAA 486

```

```
seq_name: gb_in:HAY12271
```

```

seq_documentation_block:
LOCUS      HAY12271                867 bp    mRNA    linear    INV 29-MAY-1998
DEFINITION H.armigerma mRNA for putative serine protease, clone HAT57.
ACCESSION  Y12271
VERSION     Y12271.1  GI:2463059
KEYWORDS   putative; serine protease; trypsin-like protease.
SOURCE      cotton bollworm.
ORGANISM   Helicoverpa armigera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
            Ditrysia; Noctuidae; Noctuidae; Heliothinae; Helicoverpa.
            1 (bases 1 to 867)
REFERENCE  1
AUTHORS    Bown,D.P., Wilkinson,H.S. and Gatehouse,J.A.
            Differentially regulated inhibitor-sensitive and insensitive
            protease genes from the phytophagous insect pest, Helicoverpa
            armigera, are members of complex multigene families
            Insect Biochem. Mol. Biol. 27 (7), 625-638 (1997)
JOURNAL     MEDLINE
MEDLINE     98067794
REFERENCE  2
AUTHORS     2 (bases 1 to 867)
TITLE       Direct Submission
            Submitted (26-MAR-1997) D.P. Bown, University Of Durham, Biological
            Sciences, Science Laboratories, South Road, Durham, DH1 3LE, UK
FEATURES
source
1..867
/organism="Helicoverpa armigera"
/db_xref="taxon:29058"
/clone="HAT57"
/tissue_type="midgut"
/dev_stage="fourth instar larvae"
8..772
/codon_start=1
/product="trypsin-like protease"
/protein_id="CAA72950.1"
/db_xref="GI:2463060"
/db_xref="SPTREMBL:O18436"
/translation="MRLIALVALCAFAVAAPSNPQRIYGVSVTTIDYPTIAALLYS
NMLSTWQACGGLILNNRAILPAHCTAGDANNRIRIGSTWANGSGVNHIMNIV
HPSYNSRTMDNDIAVLRSAFTFSFNQVAAISAGIADNADVAAMAGSGTSSCG
SSSBLRHYQVLTINONTCCRNNAVATFGIATIDNMICSGMPNNGRQCGQDSGPLYHN
GIYGVCSGICGAQAFPGVNAARSRTYSISSNA"

```

```

BASE COUNT      207 a      268 c      189 g      203 t
ORIGIN

```

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x HAY12271 ..

Align seg 1/1 to: HAY12271 from: 1 to: 867

13 Leuleuthralaleuvalthr 20

207 CTGCTCACTGCTGCTGTGAGC 230

seq_name: gb_in:HAY12269

seq_documentation_block:

LOCUS HAY12269 870 bp mRNA linear INV 29-MAY-1998
DEFINITION H. armigera mRNA for putative serine protease, clone HATC11, HATC11,
HAT54, SR99.
ACCESSION Y12269.1 GI:2463055
KEYWORDS putative; serine protease; trypsin-like protease;
SOURCE cotton bollworm
ORGANISM Helicoverpa armigera

REFERENCE

1. Pterygota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Diptera; Neoptera; Endopterygota; Lepidoptera; Glossata;
Ditrysia; Noctuoidea; Noctuidae; Heliothinae; Helicoverpa.
2. (bases 1 to 870)
Bowen, D.P., Wilkinson, H.S. and Gatehouse, J.A.
Differentially regulated inhibitor-sensitive and insensitive
protease genes from the phytophagous insect pest, Helicoverpa
armigera, are members of complex multigene families
Insect Biochem. Mol. Biol. 27 (7), 625-638 (1997)

AUTHORS

TITLE

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (26-MAR-1997) D.P. Bowen, University Of Durham, Biological
Sciences, Science Laboratories, South Road, Durham, DH1 1LE, UK

Location/Qualifiers
1. 870
/organism="Helicoverpa armigera"
/db_xref="taxon:29058"
/clone="SR99"
/clone="HATC11"
/clone="HATC21"
/clone="HAT54"
/tissue_type="midgut"
/dev_stage="fourth instar larvae"
12. 776
/codon_start=1
/product="trypsin-like protease"
/protein_id="CAI72948.1"
/db_xref="GI:2463056"
/db_xref="SPTREMBL:O18434"
/translation="MRITATVLCFAVAVPSRIVGSGVTTTIOPTTALIS
WNLSTYMOAGGTLINRAITLTAHCTAGANNRMRTRLSITWANSGVVHNLNINII
HPSYNSRTMDNDIVLVSATFTSFNNOVRAISAGVNIADNOAWAAGCTSSGG
SSPOLRVLQVLTINONTKRNATRTGATITDMLCSGMNGRDCOCGSGSGLVHN
GIVAGVSFGICGAQAAPPGVNAVSRYTSMISNA"
81. 773
/product="trypsin-like protease"
BASE COUNT 208 a 265 c 191 g 206 t

mat_peptide
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x HAY12269 ..

Align seg 1/1 to: HAY12269 from: 1 to: 870

13 Leuleuthralaleuvalthr 20

211 CTGCTCACTGCTGCTGTGAGC 234

seq_name: gb_pat:AR020714

seq_documentation_block:

LOCUS AR020714 885 bp DNA linear PAT 05-DEC-1998
DEFINITION Sequence 3 from patent US 5789193.
ACCESSION AR020714
VERSION AR020714.1 GI:3975329
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 885)
AUTHORS Keranen, S., Aalto, M., Outola, M., Ronne, H. and Penttila, M.
TITLE Increased production of secreted proteins by recombinant eukaryotic
cells
JOURNAL Patent: US 5789193-A 3 04-AUG-1998;
FEATURES Location/Qualifiers
1. 885
/organism="unknown"
BASE COUNT 320 a 171 c 196 g 198 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AR020714/rev ..

Align seg 1/1 to reverse of: AR020714 from: 1 to: 885

12 Alaleuuthralaleuval 19

779 GCACCTTAACGGCCTTGTGTG 756

seq_name: gb_sts:CNS06FCD

seq_documentation_block:

LOCUS CNS06FCD 972 bp DNA linear STS 10-JAN-2001
DEFINITION T7 end of clone AR0AA029A04 of library AR0AA from strain CBS 732 of
Zygosaccharomyces rouxii, sequence tagged site.
ACCESSION AL396227
VERSION AL396227.1 GI:12148185
KEYWORDS STS.
SOURCE Zygosaccharomyces rouxii.
ORGANISM Zygosaccharomyces rouxii.
REFERENCE 1 (bases 1 to 972)
AUTHORS de Montigny, J., Straub, M., Potier, S., Tekala, F., Dujon, B.,
Wincker, P., Artiguenave, F. and Souclet, J.
TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 8.
JOURNAL Zygosaccharomyces rouxii(1)
PUBMED FEBS Lett. 467 (1), 52-55 (2000)
1152883
2 (bases 1 to 972)

REFERENCE

AUTHORS

Souclet, J., L., Aigle, M., Artiguenave, F., Blandin, G.,
de Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Florent, B.,
Malpertuy, A., Neuvéglise, C., Ozier, K., Papadimitrakopoulos, O., Potier, S.,
Saurin, M., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenbach, J.

```

TITLE
Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
yeast species for molecular evolution studies(1)
JOURNAL
YEBS Lett. 487 (1), 3-12 (2000)
PUBMED
11152876
REFERENCE
3 (bases 1 to 972)
AUTHORS
Genoscope.
TITLE
Direct Submmission
JOURNAL
Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT
This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
location/Qualifiers
SOURCE
1..972
/organism="Zygosaccharomyces rouxii"
/strain="CBS 732"
/db_xref="taxon:4956"
/clone="AR0AA029A04"
/clone_1lb="AR0AA"
/note="end : T7"
STS
1..972
BASE COUNT
345 a 166 c 240 g 218 t 3 others
ORIGIN
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x CNS06FCD/rev ..
Align seg 1/1 to reverse of: CNS06FCD from: 1 to: 972
86 LeuserLeuileiIeserPhelou 93
|||||
213 CTTTCCTGATATCTCTTTCTTG 190
seq_name: em_htgo_inv:AC043561
seq_documentation_block:
ID AC043561 standard; DNA; HTG; 1005 BP.
AC AC043561;
XX
XX AC043561.1
XX
XX AC043561.1
XX
DT 12-APR-2000 (Rel. 63, Created)
DT 12-APR-2000 (Rel. 63, Last updated, Version 1)
XX
XX
DE Giardia intestinalis clone EJ5353 strain WB-C6, LOW-PASS SEQUENCE SAMPLING
XX
XX HTG; HTGS-PHASED.
XX
OS Giardia intestinalis
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
XX
XX
XX [1]
XX
XX Morrison H.G., McArthur A.G., Nixon J., Eakin N.O., Kim U., Crocker M.C.,
XX Hinkle G., Holder M.E., Sogin M.L.;
XX "Giardia: a model for ancient eukaryotic genome analysis";
XX unpublished.
XX

```

XX	1	1005	contig of 1005 bp in length.
CC	*		be preserved.
CC	*		will be sequenced to completion. In the event that
CC	*		the record is updated, the accession number will
CC	*		overlap relationships among clones to be deduced.
CC	*		However, it should not be assumed that this clone
CC	*		and the order in which they appear is completely
CC	*		arbitrary. Low-pass sequence sampling is useful for
CC	*		identifying clones that may be gene-rich and allows
CC	*		sequencing reads that have not been assembled into
CC	*		contigs. Runs of N are used to separate the reads
CC	*		NOTE: This record contains 1 individual
XX			
RL			Submitted (11-APR-2000) to the EMBL/GenBank/DBS databases.
RL			Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL			Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
RT			
RA			Hinkle G., Holder M.E., Sogin M.L.;
RA			Eakin N.Q., Morrison H.G., McArthur A.G., Nixon J., Kim U., Crocker M.C.,
RP			1-1005
RN			[2]

alignment_scores:	Quality:	8.00	Length:	8
Ratio:	1.000	Gaps:	0	
Percent Similarity:	100.000	Percent Identity:	100.000	
alignment_block:	US-10-048-197-2 x ACO43561	..		
Align seg 1/1	to: ACO43561	from: 1	to: 1005	
7	ArgLeuThrIleSerAlaLeuLeu 14			
74	CGCTTGACATCTCGGCTCTCTCG 97			
seq_name:	gb_pl:AY035150			
seq_documentation_block:				
LOCUS	AY035150	1038 bp	mRNA	linear PLN 10-JUN-2001
DEFINITION	Arabidopsis thaliana putative myo-inositol monophosphatase (F13E7.19/AT3g02870)	mRNA, complete cds.		
ACCESSION	AY035150			
VERSION	AY035150.1	GI:14334953		
KEYWORDS	FLI-CDNA.			
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 1038)			
REFERENCE	Yanada,K., Ito,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D., Lee,J.M., Quach,H.L., Tang,C., Tortum,M., Yu,G., Bowers,L., Gerniocl,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Koessme,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M., Natuska,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinozaki,K., Shinozaki,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and Thellogis,A.			
AUTHORS	Full length cDNA of gene F13E7.19/AT3g02870 (GI:6728975) unpublished			
TITLE	2 (bases 1 to 1038)			

AUTHORS

Yamada, K., Liu, S. X., Sakano, H., Pham, P. K., Banh, J., Chung, M. K., Goldsmith, A. D., Lee, J. M., Quach, H. L., Tang, C. C., Toriumi, M., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamliya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M. C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C. J., Sakurai, T., Saitou, M., Seki, M., Shin, P., Southwick, A., Shinozaki, K., Davis, R. W., Ecker, J. R. and Theologis, A.

Submitted (10-MAY-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-Length cDNA'): Seki, M., Narusaka, M., Ishida, J., Saitou, M., Kamliya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

COMMENT

TITLE
JOURNAL

The Sak, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S. X., Sakano, H., Pham, P. K., Banh, J., Chung, M. K., Goldsmith, A. D., Lee, J. M., Quach, H. L., Tang, C. C., Toriumi, M., Yu, G., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M. C., Miranda, M., Nguyen, M., Palm, C. J., Shin, P., Southwick, A., Davis, R. W., Ecker, J. R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

FEATURES

SOURCE

1. 1038
location/Qualifiers
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="3"
/clone="RAFL04-13-N21 (R09692)"
/note="This clone is in a modified pBluescript vector (Lambda ZAP) as a XhoI/SstI insert."
ecotype: Columbia"
1. 1038
/gene="F13E7_19/AT3g02870"
1. 88
/gene="F13E7_19/AT3g02870"
89. 904
/gene="F13E7_19/AT3g02870"
/codon_start=1
/evidence=experimental
/product="putative myo-inositol monophosphatase"
/protein_id="AK59654.1"
/db_xref="GI:14334954"
/translation="MADNSLDQFLAAIDAAKAGQIIIRKGFETKHEHKGQVDIV
TETDKCEELVFNHKLQLEPNHKEFGEETAFGVTELTDEPTWIVDPLDGTTPVHS
FPFVCSIGLITGKVPVGVVYVPIIMEELFTGVGGKFGINKRKVSAQSEELTAL
VTEAGTRKRDATIDPTNRINSLTKVSTIRNSGSCALDIDGVCGRVDIYELEGFGS
PMDIAGIVIVKEAGGLIFDPSGRKIDITTSORIASNSLKELEFAELRLNGA"
905. 1038
/gene="F13E7_19/AT3g02870"
3' UTR
BASE COUNT 312 a 202 c 234 g 290 t
ORIGIN

3' UTR
BASE COUNT
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AY035150

Align seg 1/1 to: AY035150 from: 1 to: 1038

13 LeuleuThralaleuLeuValThr 20
|||||
551 CTTTAAACCGCTTGTCTCGTGACA 574

seq_name: em_hltgo_inv:AC046945

seq_documentation_block:

ID AC046945 standard; DNA; HTG; 1086 bp.

AC046945;

AC046945.1

14-APR-2000 (Rel. 63, Created)

14-APR-2000 (Rel. 63, Last updated, Version 1)

Giardia intestinalis clone KI0649 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.

HTG; HTGS_PHASE0.

Giardia intestinalis

Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.

[1]

1-1086
Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,

Hinkle G., Holder M.E., Sogin M.L.;
"Giardia: a model for ancient eukaryotic genome analysis";

unpublished.

[2]

1-1086
Kim U., Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Crocker M.C.,

Hinkle G., Holder M.E., Sogin M.L.;

Submitted (13-APR-2000) to the EMBL/GenBank/DBJ databases.

Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,

Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA

* NOTE: This record contains 1 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

1 1086: contig of 1086 bp in length.

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AC046945/rev

Align seg 1/1 to reverse of: AC046945 from: 1 to: 1086

7 ArgpLeuThrIleSerAlaLeuLeu 14
|||||
981 CGCTTGACAACTCTGCTCTTCTG 958

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

OM of: us-10-048-197-2 to: EST:* out_format : pfs

Date: Sep 18, 2002 9:07 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame-r2n.model -DEV=xlp
-O=gen2_1/OSPTO.spool/US10048197/runat_17092002.135956-23526/app_query.fasta_1.168
-DB=EST -QEXT=fastap -SUFFIX=oli.rst -GAPOP=4.500 -GAPEXT=0.050
-MINMATCH=0.050 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -GAPOP=60.000 -GAPEXT=60.000 -GAPOP=6.000
-DELXT=7.000 -START=1 -MATRIX=oli -TRANS=human40.cdi -LIST=45
-DOCLIN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=40
-MODE=LOCAL -OUTFMT=pfs -NOR=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10048197 -CGCN1_1.4116 -NCPU=6
-ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-10-048-197-2

Query length: 111

Database: EST*

Database sequences: 13736207

Search time (sec): 1801.620000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb_est1:AA563389	+	10.00	169.08	3.59	464 AA563389 V175C02.r1 Knowles Sol
gb_gss:AA148814	+	10.00	150.15	5.40	672 AG148814 Pan troglodytes DNA, C
gb_est1:AA1049267	+	9.00	150.15	40.70	517 AT049267 uc85e06.y1 Sugano mus
gb_est2:BA1655178	+	9.00	146.76	62.92	768 B1655178 603280864.r1 NIH_CGAP_M
gb_gss:BA473329	+	9.00	146.67	63.64	776 BA473329 B06F02F B06 Brassic
gb_est2:BB67654	+	8.00	145.67	72.35	107 BE67654 d15le06.y1 Xenopus lae
gb_gss:AA567757	-	8.00	140.26	144.79	201 AA567757 238V604 PV MBN #26/#2
gb_est1:AA470169	-	8.00	140.26	144.79	201 AA470169 1M024812F Mouse 10kb
gb_est1:AA4731944	-	8.00	139.92	151.14	202 AA4731944 AV31944 RIKEN full-16
gb_est2:BB826421	+	8.00	139.88	151.94	210 BB826421 QV4-EN0040-25500-227-
gb_est1:AA1231241	+	8.00	139.63	151.94	210 AA1231241 EST127929 Normalized
gb_gss:BA530215	-	8.00	139.56	158.32	216 BA530215 B0HBI50TF B0H Brassic
gb_est2:BA093812	-	8.00	138.63	178.41	243 BA093812 ne6d08.s1 NCI_CGAP_KI
gb_est1:AA480791	-	8.00	137.45	207.70	279 AA480791 2M0022601R Mouse 10kb
gb_gss:AA784705	-	8.00	137.03	210.16	282 AA784705 2M0022612R Mouse 10kb
gb_est1:AA631270	-	8.00	137.03	210.16	293 AA631270 AV631270 Chlamydomonas
gb_gss:AA148390	+	8.00	136.79	223.79	301 AA148390 x11le07.x1 NCI_CGAP_KI
gb_est1:BA477818	+	8.00	136.79	223.79	301 BA477818 BA477818 RIKEN full-16
gb_est2:BA029356	+	8.00	136.65	229.92	306 BA029356 B0J93812 NIBB Mochii R
gb_est1:AA582559	+	8.00	136.38	238.20	316 AA582559 AT582559 LTI_NF010.BC
gb_est2:BB648930	+	8.00	135.70	259.86	342 BB648930 UT-M-BH2.3-ae1-f-07-0-
gb_est2:BB513403	-	8.00	135.65	261.55	344 BB513403 dae04c10.x1 NICHG XCC
gb_gss:BA113823	-	8.00	135.43	269.07	353 BA113823 TM-ad42B03-SKPL Trich
gb_est2:BA277543	-	8.00	135.31	273.27	358 BA277543 RPCI-24-252K17.TV RPCI
gb_gss:BA033149	+	8.00	135.12	279.99	366 BA033149 RPCI-24-294H20.TV RPCI
gb_est1:AA720873	+	8.00	135.07	281.68	368 AA720873 UT-M-BH2.3-ae1-f-07-0-
gb_gss:AA301765	+	8.00	134.89	288.42	371 AA301765 AF-F4F Resistance gene
gb_gss:AA649818	+	8.00	134.89	288.42	376 AA649818 1M051908R Mouse 10kb
gb_est1:BA816077	+	8.00	134.66	296.03	385 BA816077 BB816077 RIKEN full-16
gb_est1:AA720873	+	8.00	134.53	296.88	386 AA720873 TM-ad42B03-SKPL Trich
gb_est2:BA065087	-	8.00	134.42	306.20	392 BA065087 OB56905.s1 NCI_CGAP_KI
gb_gss:AA008615	+	8.00	134.26	313.85	406 AA008615 nbe016097f D02_MG Rice
gb_est2:BA039807	+	8.00	134.14	317.25	410 BA039807 HS_2208_B1-D02_MG CRT
gb_est2:BA072077	+	8.00	134.12	318.10	411 BA072077 B0J93812 NIBB Mochii R

gb_gss:AA132522 + 8.00 133.79 331.76 427 | AA132522 OSJNB0070H1f CUGI
gb_est2:BB685620 - 8.00 133.73 334.32 430 | BB685620 u56d10.y1 Soares t
gb_est1:BB848843 + 8.00 133.65 337.75 438 | BB848843 BB848843 RIKEN full
gb_gss:BB6868 + 8.00 133.58 341.17 438 | BB6868 RPCI11-26H10.TV RPCI
gb_gss:AA313383 + 8.00 133.46 346.32 444 | AA313383 RPCI11-102020.TV RP
gb_gss:AA301770 - 8.00 133.44 347.18 445 | AA301770 AF-E8R Resistance g
gb_gss:AA0406823 - 8.00 133.44 347.18 445 | AA0406823 HS_5092_B1_B04_Sp6E

seq_name: gb_est1:AA563389

seq_documentation_block:

LOCUS AA563389 464 bp mRNA linear EST 18-AUG-1997
DEFINITION V175C02.r1 Knowles Solter mouse blastocyst B1 Mus musculus CDNA
clone IMAGE:978050 5', mRNA sequence.

ACCESSION AA563389
VERSION AA563389.1 GI:2334854
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 464)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:558778
Putative full length read
vector to vector length is 512.
Location/Qualifiers
1..464
/organism="Mus musculus"
/strain="B6D2 F1/3T"
/db_xref="taxon:10090"
/clone="IMAGE:978050"
/clone_lib="Knowles Solter mouse blastocyst B1"
/tissue_type="Placental"
/dev_stage="embryo (pre-implantation)"
/lab_note="DH10B"
/note="Organ: embryo; Vector: pSPORT; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(GT):
5'-CGGTCACGCGACGCGTGTGTGTGT-3'. cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (larger
inserts) and B3."

FEATURES

source

BASE COUNT 97 a 109 c 110 g 148 t
ORIGIN
alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x AA563389 ..
Align seg 1/1 to: AA563389 from: 1 to: 464
10 HleeralaleuethrAlaaleuVal 19

```

|||||
329 ATCAGTCTCTGCTAACGCCCTGCTGCTC 358
seq_name: gb_gss:AG148814

seq_documentation_block:
LOCUS      AG148814                672 bp    DNA        linear    GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-011C03.TJ, genomic survey
sequence.
ACCESSION  AG148814
VERSION    AG148814.1 GI:16678492
KEYWORDS   GSS: GSS (genome survey sequence).
SOURCE     Pan troglodytes male lymphocytes DNA, clone: RP43-011C03.TJ.
ORGANISM   Male BAC library clone:RP43-011C03.TJ.
            Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE  1 (sites)
            Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Toki,Y., Watanabe,H. and Sakaki,Y.
            BAC end sequences of library RPCI-43
            Unpublished
            2 (bases 1 to 672)
            Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Toki,Y., Watanabe,H. and Sakaki,Y.
            Direct Submission
            Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            E-mail:chimbse@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
            Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
            PRIMERS
            Sequencing: TJ
            LIBRARY
            Vector : pBACe3.6
            R.Site 1 : EcoRI
            R.Site 2 : EcoRI.
FEATURES
    source
        1..672
            location/Qualifiers
            /organism="Pan troglodytes"
            /db_xref="taxon:9598"
            /clone="RP43-011C03.TJ"
            /sex="male"
            /cell_type="lymphocytes"
            /clone_lib="RPCI-43 Chimpanzee Male BAC library"
BASE COUNT      237 a      87 c      99 g      249 t
ORIGIN
..
alignment_scores:
    Quality:      10.00      Length:      10
    Ratio:        1.000      Gaps:      0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x AG148814 ..
Align seg 1/1 to: AG148814 from: 1 to: 672
86 LeuSerLeuIleIleSerPheLeuTPVal 95
|||||
130 CTTTCATTATATATCATTTCTTGGTA 159
seq_name: gb_est1:AI049267

seq_documentation_block:
LOCUS      AI049267                517 bp    mRNA        linear    EST 08-JUL-1998
DEFINITION uc85e06.y1 Sugano mouse kidney mklia Mus musculus cDNA clone
IMAGE:14322450 5', mRNA sequence.
ACCESSION  AI049267

```

```

VERSION      AI049267.1 GI:3297554
KEYWORDS     house mouse.
SOURCE       Mus musculus
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
REFERENCE    1 (bases 1 to 517)
            Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
            The WashU-HMI Mouse EST Project
            Unpublished (1996)
            Contact: Marra M/Mouse EST Project
            WashU-HMI Mouse EST Project
            Washington University School of Medicine
            444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watsn.wustl.edu
            This clone is available royalty-free through LML; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:916518
            Seq primer: custom primer used
            High quality sequence stop: 462.
FEATURES
    source
        1..517
            location/Qualifiers
            /organism="Mus musculus"
            /strain="C57BL"
            /db_xref="taxon:10090"
            /clone="IMAGE:1432450"
            /clone_lib="Sugano mouse kidney mklia"
            /sex="female"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
            (CACTGATG); Site_2: DraIII (CACTCATG); 1st strand cDNA
            was primed with an oligo(dT) primer
            [ATGTGGCCCTTTTCTTTT]; double-stranded cDNA was
            ligated to a DraIII adaptor [TGTGGCCCTACTG]; digested
            and cloned into distinct DraIII sites of the pME18S-FL3
            vector (5' site CACTGATG, 3' site CACTCATG); xhoI should
            be used to isolate the cDNA insert. Size selection was
            performed to exclude fragments <1.5kb. Library
            constructed by Dr. Sumio Sugano (University of Tokyo
            Institute of Medical Science). Custom primers for
            sequencing: 5' end primer CTTGCTCTCTAAAGCTGCG and 3' end
            primer CGACCTGACGCTGACGACA."
BASE COUNT      108 a      147 c      152 g      110 t
ORIGIN
..
alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:        1.000      Gaps:      0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x AI049267 ..
Align seg 1/1 to: AI049267 from: 1 to: 517
13 LeuLeuThraIaLeuLeuValTrngI 21
|||||
30 CTGCTTACAGCCCTCCTGTGACGGG 56
seq_name: gb_est2:BI655178

seq_documentation_block:
LOCUS      BI655178                768 bp    mRNA        linear    EST 12-SEP-2001
DEFINITION 603280984P1 NIH_GCAP_Mam4 Mus musculus cDNA clone IMAGE:5325357 5',
mRNA sequence.

```

ACCESSION B1655178
 VERSION B1655178.1 GI:15569414
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 768)
 NIH-MGC <http://mgc.ncl.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L1AM1825 row: g column: 22
 High quality sequence stop: 760.
 Location/Qualifiers
 1..768
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone IMAGE:5325357
 /clone_1lb="NIH-CGAP_Mam4"
 /tissue_type="tumor, gross tissue"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: PCMV-SPORE; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert 2.5 kb. Library constructed by Life Technologies, catalog # 12018-016. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996). Note: this is a NCI-CGAP Library."
 BASE COUNT 180 a 214 c 225 g 149 t
 ORIGIN
 alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 Alignment block:
 US-10-048-197-2 x B1655178 ..
 Align seg 1/1 to: B1655178 from: 1 to: 768
 8 LeuThrlleSeRaLaLeuThraLa 16
 ||||||||||||||||||||
 180 CTCGACATCTCCGCTCTTACAGCA 206
 seq_name: gb_gss:BH473529
 seq_documentation_block:
 LOCUS BH473529 776 bp DNA linear GSS 13-DEC-2001
 DEFINITION BCGDF027F BCGD Brassica oleracea genomic clone BCGDF02, DNA
 sequence.
 ACCESSION BH473529
 VERSION BH473529.1 GI:17681640
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 776)
 REFERENCE
 1 Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea
 JOURNAL Unpublished (2001)
 COMMENT Other GSSs: BCGDF027F
 CONTACT: Chris Town
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.
 Location/Qualifiers
 1..776
 /organism="Brassica oleracea"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BCGDF02"
 /clone_1lb="BCGD"
 /note="Vector: PHOS1; Site:1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"
 BASE COUNT 197 a 142 c 218 g 219 t
 ORIGIN
 alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 Alignment block:
 US-10-048-197-2 x BH473529 ..
 Align seg 1/1 to: BH473529 from: 1 to: 776
 99 ArgProLysAsnLeuSerPheTyrLeu 107
 ||||||||||||||||||||
 502 AGACCTAAATAATCTACTTTTATCTC 528
 seq_name: gb_est2:BE679654
 seq_documentation_block:
 LOCUS BE679654 107 bp mRNA linear EST 18-APR-2001
 DEFINITION df51e06.y1 Xenopus laevis unfertilized egg cDNA library Xenopus
 laevis cDNA clone IMAGE:3742882 5', mRNA sequence.
 ACCESSION BE679654
 VERSION BE679654.1 GI:10062553
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.
 1 (bases 1 to 107)
 Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood, R., Theising, B., Bowers, Y., Peterson, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R. and Wilson, R.
 Washu Xenopus EST project, 1999
 Unpublished (1999)
 CONTACT: Sandy Clifton, Ph.D.
 Washu Xenopus EST project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 Library constructed by Bruce Blumberg
 Library normalized by Jihwan Song
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clone distribution information for this library can be found through Research Genetics, visit their web page at: <http://www.resgen.com/> Please reference the id listed

below when ordering this clone: Source lab clone id - xlmeg001g11
Seq primer: -40RP from Gibco
High quality sequence stop: 83.
Location/Qualifiers

FEATURES
SOURCE

1. 107
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3742882"
/clone_1lb="Xenopus laevis unfertilized egg cDNA library"
/tissue_type="unfertilized egg"
/lab_host="Top-10 F."

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from 2ug of poly A+ RNA. EcoRI-XhoI cut cDNA was then ligated into UniZap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. SS-library phagmids were prepared by mass excision from the original library and normalized by hybridization to biotinylated driver (prepared from the same library by PCR) to Cot-omega of 11. After removal of hybrids and excess driver by streptavidin sepharose chromatography, the ss-phagmids were made double stranded and electroporated into Top-10 F+. Original library construction by Bruce Blumberg (Blumberg et al., 1991 Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9, 2923-2935). Note: This is a Xenopus Gene Collection (XGC) library."

BASE COUNT 32 a 17 c 36 g 21 t 1 others
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x BE679654/rev ..

Align seg 1/1 to reverse of: BE679654 from: 1 to: 107

11 SerAlaLeuThrAlaLeu 18

|||||
90 TCTGCTCTCTCACACGCTG 67

seq_name: gb_gss:A2567757

seq_documentation_block:

LOCUS A2567757 198 bp DNA linear GSS 07-MAY-2001
DEFINITION 238PVG04 PV MBN #26/#27 (amplified once) Plasmodium vivax genomic
3', DNA sequence.

ACCESSION

A2567757

VERSION A2567757.1 GI:13976175

KEYWORDS

GSS.

SOURCE

malaria parasite P. vivax.

ORGANISM

Plasmodium vivax

REFERENCE

1 (bases 1 to 198)

AUTHORS

Carlton,J.M.-R. and Dame,J.B.

TITLE

The Plasmodium vivax and P. berghei gene sequence tag projects

JOURNAL

Parasitol. Today 16 (10), 409 (2000)

COMMENT

Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: M13(-20) forward
Class: shotgun.

FEATURES

source

Location/Qualifiers
1. 198
/organism="Plasmodium vivax"
/strain="Belém"

/db_xref="taxon:5855"
/clone_1lb="PV MBN #26/#27 (amplified once)"
/dev_stage="asexual blood forms"
/lab_host="Saimiri boliviensis"

/note="Vector: Lambda ZAP II (Stratagene); individual clones excised into phagemid pBluescript; Site_1: EcoRI; Site_2: EcoR I; Genomic DNA was prepared from asynchronous blood stage forms of the Belém line of P. vivax grown in squirrel monkeys. Parasitized erythrocytes were purified from contaminating host leukocytes by filtration of ADP activated blood through acid-washed glass beads and Whatman CPl cellulose columns by gravity filtration. Purified DNA was digested with mung bean nuclease in the presence of 42.5% formamide at 50°C as described (Gallinski et al., 1988, N.A.R. 16, 6883-6896). Eco RI linkers were added and the constructs ligated into Lambda ZAP II. P. vivax Belém was originally isolated from a patient in Belém, Brazil 1980 by Mercia de Arruda, adapted to Saimiri monkeys by Jurgen Gysin, and maintained since 1983 in squirrel monkeys."

BASE COUNT 56 a 43 c 62 g 37 t
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x A2567757/rev ..

Align seg 1/1 to reverse of: A2567757 from: 1 to: 198

11 SerAlaLeuThrAlaLeu 18

|||||
189 TCTGCGCTTTGACTGCCTCTG 166

seq_name: gb_gss:A2470169

seq_documentation_block:

LOCUS A2470169 201 bp DNA linear GSS 04-OCT-2000
DEFINITION IM0284B12F Mouse 10kb plasmid UGCG1M library Mus musculus genomic
clone UGCG1M0284B12 F, DNA sequence.

ACCESSION

A2470169

VERSION A2470169.1 GI:10628294

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 201)

AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weis,R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SUC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0284 row: B column: 12
Seq primer: CGTGTAAACGACGCCACG
Class: plasmid ends
High quality sequence stop: 201.

```

FEATURES
  source
      1. 201
         Location/Qualifiers
           /organism="Mus musculus"
           /strain="C57BL/6J"
           /db_xref="taxon:10090"
           /clone="U081M0284812"
           /clone_11b="Mouse 10kb plasmid U081M 11b"
           /sex="Male"
           /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
           /note="Vector: PMD29v; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
           (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g1473214|9b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT      38 a      73 c      50 g      40 t
ORIGIN

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x A2470169/rev ..

Align seg 1/1 to reverse of: A2470169 from: 1 to: 201

11 SerAlaLeuThrAlaLeu 18
|||||
72 TCTGCACTGCTGACTGCCCTGCTG 49

seq_name: gb_est1:AV331944

seq_documentation_block:
LOCUS      AV331944      202 bp      mRNA      linear      EST 11-NOV-1999
DEFINITION AV331944 RIKEN full-length enriched, adult male medulla oblongata
ACCESSION  AV331944
VERSION     AV331944.1 GI:6371996
KEYWORDS   EST.
SOURCE      house mouse.
ORGANISM   Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1 (bases 1 to 202)
AUTHORS     Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishi,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,T., Kai,
            C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,
            Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateo,M., Tomioka,N., Tsunoda,Y.,
            Watabiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
            RIKEN Mouse ESTs (Konno,H., et al. 1999)
COMMENT     Contact: Yoshihide Hayashizaki

```

```

FEATURES
  source
      1. 202
         Location/Qualifiers
           /organism="Mus musculus"
           /strain="C57BL/6J"
           /db_xref="taxon:10090"
           /clone="630528C24"
           /clone_11b="RIKEN full-length enriched, adult male medulla oblongata"
           /sex="male"
           /tissue_type="medulla oblongata"
           /dev_stage="adult"
           /lab_host="DH10B"
           /note="Site 1: Salt; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTATTTTATN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAATTAATTAATCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: Salt; 3' end: BamHI"

BASE COUNT      101 a      20 c      51 g      30 t
ORIGIN

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AV331944/rev ..

Align seg 1/1 to reverse of: AV331944 from: 1 to: 202

103 LeuSerPheTyrLeuThrAlaIys 110
|||||
100 CTTTCCTTCTATCTTACGACCA 77

seq_name: gb_est2:BE826421

```

```

seq_documentation_block:
LOCUS      BE826421                209 bp    mRNA    linear    EST 22-SEP-2000
DEFINITION QV4-EN0040-250500-227-cl1 EN0040 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BE826421
VERSION    BE826421.1  GI:10256799
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 209)
AUTHORS   Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
            Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
            Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,
            Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
            M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
            Simpson, A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil.
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LIRC Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?l=6t2-QV4-EN0040-250
            500-227-cl1&t3=2000-05-25&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 87
            High quality sequence stop: 209.
FEATURES
    source
        1..209
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="EN0040"
            /dev_stage="Adult"
            /note="Organ: lung; normal; Vector: puc18; Site_1: Sma1;
            Site_2: Sma1; A mini-library was made by cloning products
            derived from ORESTES PCR (U.S. Letters Patent application
            No. 196,716 - Ludwig Institute for Cancer Research)
            tissue into the puc 18 vector. Reverse transcription of
            tissue mRNA and cDNA amplification were performed under
            low stringency conditions."
BASE COUNT      52 a          64 c          53 g          40 t
ORIGIN

```

```

alignment_scores:
    Quality:      8.00      Length:      8
    Ratio:        1.000     Gaps:        0
    Percent Similarity: 100.000    Percent Identity: 100.000

```

alignment_block:

```

US-10-048-197-2 x BE826421  ..
Align seg 1/1  to: BE826421  from: 1  to: 209

```

```

7 ArgLeuThrIleSerAlaLeuLeu 14
|||||
166 CGGCTTACCATCTCGGCTCTCTC 189

```

seq_name: gb_est1:AI231241

seq_documentation_block:

```

LOCUS      AI231241                210 bp    mRNA    linear    EST 31-JAN-1999
DEFINITION EST227929 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
            REMDHS3 3', end, mRNA sequence.

```

```

ACCESSION  AI231241
VERSION    AI231241.1  GI:3815121
KEYWORDS   EST.
SOURCE     Rattus sp.
ORGANISM   Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 210)
AUTHORS   Lee, N.H., Glodet, A., Chandra, I., Mason, T.M., Quackenbush, J.,
            Kerlavage, A.R. and Adams, M.D.
TITLE      Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
            Gene Index
JOURNAL    Unpublished (1998)
COMMENT    The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@ligr.org
            Seq primer: M13-21.
FEATURES
    source
        1..210
            /organism="Rattus sp."
            /db_xref="ATCC (lnost):2037375"
            /db_xref="taxon:10118"
            /clone="REMDHS3"
            /clone_lib="Normalized rat embryo, Bento Soares"
            /dev_stage="embryo 8, 12, 18 dpc"
            /note="Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT      51 a          42 c          37 g          80 t
ORIGIN

```

```

alignment_scores:
    Quality:      8.00      Length:      8
    Ratio:        1.000     Gaps:        0
    Percent Similarity: 100.000    Percent Identity: 100.000

```

alignment_block:

```

US-10-048-197-2 x AI231241  ..
Align seg 1/1  to: AI231241  from: 1  to: 210

```

```

85 IleuSerLeuIleIleSerPhe 92
|||||
85 ATCTTCTCTGATATCTCTTT 108

```

seq_name: gb_gss:BH530215

seq_documentation_block:

```

LOCUS      BH530215                216 bp    DNA    linear    GSS 14-DEC-2001
DEFINITION BOHB150TF BOHB Brassica oleracea genomic clone BOHB150, DNA
            sequence.
ACCESSION  BH530215
VERSION    BH530215.1  GI:17746359
KEYWORDS   GSS.
SOURCE     Brassica oleracea.
ORGANISM   Brassica oleracea.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

```

```

REFERENCE  1 (bases 1 to 216)
AUTHORS   Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.
TITLE      Whole genome shotgun sequencing of Brassica oleracea
JOURNAL    Unpublished (2001)
COMMENT    Other_GSSs: BOHB150TF
            Contact: Chris Town
            TIGR

```

```

9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@ligr.org

```

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF
Class: sheared ends.

FEATURES

source

Location/Qualifiers
1. 216
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOH8150"
/clone_1lb="BOH8"
/note="Vector: PHOS1; site_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 69 a 18 c 50 g 79 t
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x BH530215/rev ..

Align seg 1/1 to reverse of: BH530215 from: 1 to: 216

86 Leuserleu1le1eserphelen 93
|||||
181 CTAACTCTAATTAATTCCTTCTTA 158

seq_name: gb_est2:BJ093812

seq_documentation_block:

LOCUS BJ093812 218 bp mRNA linear EST 12-DEC-2001
DEFINITION BJ093812 NIBB Mochii normalized xenopus early gastrula library
Xenopus laevis cDNA clone XL143j12.5', mRNA sequence.

ACCESSION BJ093812
VERSION BJ093812.1 GI:17593552

KEYWORDS

EST.

SOURCE

African clawed frog.
Xenopus laevis

REFERENCE

1 (bases 1 to 218)
Kikayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-I, T. and Kohara, Y.

AUTHORS

TITLE

JOURNAL

Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadasu Shin-I
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tsuhin@genes.nig.ac.jp.

FEATURES

source

Location/Qualifiers
1. 218
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XL143j12"
/clone_1lb="NIBB Mochii normalized Xenopus early gastrula library"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"

BASE COUNT 56 a 46 c 58 g 56 t 2 others
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x BJ093812/rev ..

Align seg 1/1 to reverse of: BJ093812 from: 1 to: 218

11 Ser1aleu1leu1rhalaleu1 18
|||||
65 TCTGCTCTTCTCAGCAGCTGCTG 42

seq_name: gb_est1:AA480791

seq_documentation_block:

LOCUS AA480791 243 bp mRNA linear EST 14-NOV-1997
DEFINITION ne66d08.s1 NCI-CGAP_K1d1 Homo sapiens cDNA clone IMAGE:911151, mRNA
sequence.
ACCESSION AA480791
VERSION AA480791.1 GI:2209489
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 243)

REFERENCE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNI at:
www.bio.lnlnl.gov/bdrp/image/image.html
Insert Length: 427 Std Error: 0.00
Seq primer: 41m13 fwd. ET from Amersham.

FEATURES

source

Location/Qualifiers
1. 243
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="911151"
/clone_1lb="NCI-CGAP_K1d1"
/tissue_type="Kidney"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from invasive kidney
tumor, cDNA made by oligo-dT priming. Non-directionally
cloned. Size selected on agarose gel, average insert
size 600 bp. Reference: Kitzman et al. (1996) Cancer
Research 56:5380-5383."

BASE COUNT 98 a 56 c 50 g 39 t
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AA480791/rev ..

Align seg 1/1 to reverse of: AA480791 from: 1 to: 243

52 lleserArpghglyserala 59

|||||
131 ATTTCCTTCGCTTGGGCTGCT 108

seq_name: gb_gss:A2782323

seq_documentation_block:

LOCUS A2782323 279 bp DNA linear GSS 16-FRB-2001

DEFINITION 2M0022G01R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC2M0022G01 R, DNA sequence.

ACCESSION A2782323

VERSION A2782323.1 GI:12915930

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 279)

REFERENCE

AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

TITLE

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 1000 Std Error: 0.00

Plate: 0022 row: G column: 01

Seq primer: CACACAGGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 279.

Location/Qualifiers

1..279

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0022G01"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (g114732114[gb|AF129072.1]), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT

128 a 46 c 52 g 53 t

alignment_scores:

Quality: 8.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x A2782323/rev

Align seg 1/1 to reverse of: A2782323 from: 1 to: 279

86 LeuSerLeu1le1leSerPheLeu 93
|||||

217 CTGAGTTTGATATATTCCTTCCTT 194

seq_name: gb_gss:A2784705

seq_documentation_block:

LOCUS A2784705 282 bp DNA linear GSS 16-FRB-2001

DEFINITION 2M0027G12R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC2M0027G12 R, DNA sequence.

ACCESSION A2784705

VERSION A2784705.1 GI:12920711

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 282)

REFERENCE

AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

TITLE

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 1000 Std Error: 0.00

Plate: 0027 row: G column: 12

Seq primer: CACACAGGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 282.

Location/Qualifiers

1..282

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0027G12"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (g114732114[gb|AF129072.1]), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT

66 a 94 c 80 g 42 t

ORIGIN

alignment_scores:

Quality: 8.00

Length: 8

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AZ784705/rev ..

Align seg 1/1 to reverse of: AZ784705 from: 1 to: 282

11 SerAlaLeuThrAlaLeu 18
|||||
228 TCTGCACTGCTGCTGCTGCTG 205

seq_name: gb_est1:AV631270

seq_documentation_block:

LOCUS AV631270 293 bp mRNA linear EST 15-DEC-2000
DEFINITION AV631270 Chlamydomonas reinhardtii 5' to 0.04% CO2 Chlamydomonas
reinhardtii cDNA clone LCL091B11_r 5', mRNA sequence.

ACCESSION AV631270
VERSION AV631270.1 GI:10793904

KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.

ORGANISM

Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

REFERENCE

AUTHORS

Asamizu, E., Miura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohnaka, K.,
Nakamura, Y., and Tabata, S.
Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
DNA Res. 7 (5), 305-307 (2000)

JOURNAL

MEDLINE

Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yama 1332-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

SOURCE

1..293
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"

/strain="C9"

/db_xref="taxon:3055"

/clone_lib="LCL091B11_r"

/note="Vector: pBluescriptII SK⁺ Site_1: EcoRI; Site_2:

XhoI; The cDNA library was constructed from cells cultured

in a carbon stress acclimatized condition in which carbon

dioxide concentration in the bubbling gas was changed from

5% to 0.04%"

BASE COUNT 70 a 85 c 89 g 49 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AV631270/rev ..

Align seg 1/1 to reverse of: AV631270 from: 1 to: 293

7 ArgLeuThrIleSerAlaLeu 14
|||||
207 CGCTTAACGATTCTGCTGCTG 184

seq_name: gb_est1:AW148390

seq_documentation_block:

LOCUS AW148390 301 bp mRNA linear EST 03-NOV-1999
DEFINITION xfl107.x1 NCI-CGAP Kid8 Homo sapiens cDNA clone IMAGE:2617764.3'
similar to contains Alu repetitive element; contains element MSRI

MSRI repetitive element ;, mRNA sequence.

ACCESSION AW148390
VERSION AW148390.1 GI:6196211

KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 301)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgap@db-remail.nih.gov

Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.

Emmett-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/BLN at:

www.bio.lnl.gov/db/ftp/image.html

Seq primer: -40UP from Gibco

FEATURES

SOURCE

1..301
Location/Qualifiers
/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:2617764"

/clone_lib="NCI-CGAP Kid8"

/tissue_type="renal cell tumor"

/lab_host="DH10B"

/note="Organ: Kidney; Vector: pCMV-Sport6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.2 kb. Life Technologies catalog #:

11524-014"

BASE COUNT

53 a 90 c 76 g 80 t 2 others

ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AW148390 ..

Align seg 1/1 to: AW148390 from: 1 to: 301

9 ThrIleSerAlaLeuThrAla 16
|||||

55 ACGATCTGCTGCTTGTACTGCA 78

seq_name: gb_est1:BB477818

seq_documentation_block:

LOCUS BB477818 301 bp mRNA linear EST 23-JUL-2000
DEFINITION BB477818 RIKEN full-length enriched, 13 days embryo heart mus
musculus cDNA clone D330027A14.3', mRNA sequence.

ACCESSION BB477818
VERSION BB477818.1 GI:9395427

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 301)

Komono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci

, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,

Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,

Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,

TITLE
JOURNAL
COMMENT

Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamada, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
RIKEN Mouse ESTs (Kono, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Thermostabilization and thermocyclization of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P., and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.riken.go.jp) for further details.

FEATURES
source

Location/Qualifiers
1..301
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_id="D330027A14"
/clone_lib="RIKEN full-length enriched, 13 days embryo heart"
/tissue_type="heart"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGCGCGACACGAGTTTCTTTTCTTTTCTT 3'], cDNA was transcribed using trehalose thermo-activated reverse cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATCTCGAGTAAATTAATTAATGCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FIC I."

BASE COUNT 88 a 59 c 59 g 95 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x BB477818 ..

Align seg 1/1 to: BB477818 from: 1 to: 301

22 CysValSerThrGlyAsnValAla 29
|||||
73 TGTGTCTGACTGTAATGTGGCC 96

seq_name: gb_est2:BJ029356

seq_documentation_block:

LOCUS BJ029356 306 bp mRNA linear EST 07-DEC-2001
DEFINITION BJ029356 NIBB Mochli normalized Xenopus neurula library Xenopus laevis cDNA clone Xl012d07 5', mRNA sequence.
ACCESSION BJ029356
VERSION BJ029356.1 GI:17418919
KEYWORDS EST.
SOURCE African clawed frog.
Xenopus laevis
ORGANISM

REFERENCE
AUTHORS

Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadao Shin-i
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..306
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone_id="XL012d07"
/clone_lib="NIBB Mochli normalized Xenopus neurula library"
/tissue_type="whole embryo"
/dev_stage="stage 15"

FEATURES
source

BASE COUNT 87 a 56 c 83 g 80 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x BJ029356/rev ..

Align seg 1/1 to reverse of: BJ029356 from: 1 to: 306

11 SerAlaLeuLeuThrAlaLeuLeu 18
|||||
151 TGTGCTCTTCTCACACACTGCTG 128

seq_name: gb_est1:AL582559

seq_documentation_block:

LOCUS AL582559 316 bp mRNA linear EST 16-FEB-2001
DEFINITION AL582559 LTI_NFL010_BC2 Homo sapiens cDNA clone CSDDL010YK13 5 prime, mRNA sequence.
ACCESSION AL582559
VERSION AL582559.1 GI:12950661
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 316)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source Location/Qualifiers
1.316
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDBL010K13"
/clone_1ib="LTI_NFL010-BC2"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifestech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 63 a 101 c 77 g 68 t 7 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AL582559 ..

Align seg 1/1 to: AL582559 from: 1 to: 316

12 AlaleuLeuthrAlaleuLeuVal 19
|||||
193 GCCCTGCTAACGACGCTCTGCTGTA 216

seq_name: gb_est2:BE648930

seq_documentation_block:

LOCUS BE648930 342 bp mRNA linear EST 06-SEP-2000
DEFINITION UI-M-BH2.3-aol-f-07-0-UI.r1 NIH_BMAP_M.S3.3 Mus musculus cDNA clone
UI-M-BH2.3-aol-f-07-0-UI 5', mRNA sequence.
ACCESSION BE648930
VERSION BE648930.1 GI:9974751
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathu; Muridae; Murinae; Mus.
1 (bases 1 to 342)

AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE
COMMENT

97044477
Contact: Chln, H
National Institute of Mental Health
6001 Executive Blvd, Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@nhi.nih.gov
CDNA Library Preparation: M.B. Soares Lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements

FEATURES
source Seq primer: M13 Reverse.
Location/Qualifiers
1.342
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH2.3-aol-f-07-0-UI"
/clone_1ib="NIH_BMAP_M.S3.3"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M.S3.3 library is a subtracted library of a
series, ultimately derived from a mixture of individually
tagged normalized libraries from ten regions of the mouse
brain (cerebellum, brain stems, olfactory bulbs,
hypothalamus, cortex, amygdala, basal ganglia, pineal
gland, striatum, hippocampus) after a series of
subtractions to reduce the representation of cDNAs from
which ESTs had already been generated. The following
serially subtracted libraries were generated in this
process: NIH_BMAP_M.S3.3, NIH_BMAP_M.S2, NIH_BMAP_M.S1.
The subtracted library (NIH_BMAP_M.S3.3) was constructed
as follows: PCR amplified cDNA inserts from NIH_BMAP_M.S2
clones from which 3' ESTs had been derived was used as a
driver in a hybridization with the NIH_BMAP_M.S2 library
in the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the
NIH_BMAP_M.S3.3 library. This procedure has been
previously described (Bonaldo, Lennon and Soares, Genome
Research 6:791-806, 1996)"

BASE COUNT 72 a 71 c 76 g 123 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x BE648930 ..

Align seg 1/1 to: BE648930 from: 1 to: 342

22 CysValSerThrGlyAsnValAla 29
|||||
315 TGTGTCAGTACTGGAAATGTGCA 338

seq_name: gb_est2:BG513403

seq_documentation_block:

LOCUS BG513403 344 bp mRNA linear EST 28-MAR-2001
DEFINITION dae04c10.x1 NICHD XGC Lul Xenopus laevis cDNA clone IMAGE:4632979
3', mRNA sequence.
ACCESSION BG513403
VERSION BG513403.1 GI:13484060
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.

AUTHORS NCI-CGAP http://www.nci.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available
 through the I.M.A.G.E. Consortium/LNL at: info@image.lnl.gov
 Seq primer: -400P from Gibco.

FEATURES

source

1.344
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone_image="4632979"
 /clone_lib="NICHD XGC L1"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pCMV-Sport6; Site_1: Not;
 Site_2: Salt; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2.6 kb. Constructed by Life
 Technologies. Note: This is a Xenopus Gene Collection (XGC
) library."

BASE COUNT 118 a 70 c 84 g 72 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x BG513403/rev ..

Align seg 1/1 to reverse of: BG513403 from: 1 to: 344

10 lleserlaleu1le1serphaleu 17
 ||||||||||||||||||||
 106 ATCACTGCATCTCCTCATGCTG 83

seq_name: gb_gss:BH113823

seq_documentation_block:

LOCUS BH113823 350 bp DNA linear GSS 19-JUL-2001
 DEFINITION RPCI-24-252K17.TV RPCI-24 Mus musculus genomic clone RPCI-24-252K17
 , DNA sequence.

ACCESSION BH113823
 VERSION BH113823.1 GI:14951386

KEYWORDS

GSS
 house mouse.

SOURCE

Mus musculus

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 350)

REFERENCE
 AUTHORS Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akintet,B., Levins,M.,
 Tsugey,G., Geer,K., Krol,M., Shwartsbeyn,A., Gebregorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other_GSSs: RPCI-24-252K17.TU
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhaoc@tigr.org

TITLE

JOURNAL

COMMENT

Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@emall.cho.org). Clones may be purchased from BACPAC
 Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
 page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
 Seq primer: T7
 Class: BAC ends.

FEATURES

Location/Qualifiers

1.350
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="RPCI-24-252K17"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pPRABAC1; Site_1: BamHI; Site_2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pPRABAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 142 a 53 c 69 g 86 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x BH113823/rev ..

Align seg 1/1 to reverse of: BH113823 from: 1 to: 350

86 leuserleu1le1serphaleu 93
 ||||||||||||||||||||
 297 TTATCTGTATTTCTCTCTG 274

seq_name: gb_est2:BM277543

seq_documentation_block:

LOCUS BM277543 353 bp mRNA linear EST 20-DEC-2001
 DEFINITION Tm_ad_42B03.SKPL Trichuris muris (parasitic nematode) mixed adult
 Trichuris muris cDNA clone Tm_ad_42B03 5' similar to gb|AAC02985.21
 (AF038554), density regulated protein drpl - Homo sapiens, mRNA
 sequence.

ACCESSION BM277543
 VERSION BM277543.1 GI:117970794

KEYWORDS

EST
 Trichuris muris.

SOURCE

Trichuris muris

ORGANISM
 Eukaryota; Metazoa; Nematoda; Enoplea; Trichocephalida; Trichuridae
 ; Trichuris.
 1 (bases 1 to 353)
 REFERENCE
 AUTHORS Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Gulliano,D., Hall
 ,N., Quayle,M. and Barrell,B.
 Edinburgh University/Sanger Centre Nematode EST Project
 Unpublished (2000)
 Contact: Blaxter ML
 Institute of Cell, Animal and Population Biology
 University of Edinburgh
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
 3JT, UK.
 Tel: +44 131 650 6760
 Fax: +44 131 670 5450
 Email: mark.blaxter@ed.ac.uk

TITLE
 JOURNAL
 COMMENT
 The library was prepared by Richard Grencis, Manchester University,
 Manchester. Sequencing was performed by the Pathogen Sequencing
 Unit, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart
 Barrell).

PCR Primers
 FORWARD: T3
 BACKWARD: T7PL
 Plate: 42 row: B column: 03
 Seq primer: SKPL
 High quality sequence stop: 278.

FEATURES

source

1.353
 /organism="Trichuris muris"

```
/db_xref="taxon:70415"
/clone_fm_ad_42B03"
/clone_lib="Trichuris muris (parasitic nematode) mixed
adult"
/sex="mixed"
/dev_stage="adult"
/notes="Vector: lambda Zap II; site_1: EcoRI (5'end);
site_2: XhoI (3'end); Trichuris muris is a nematode
parasite of rodents related to the human whipworm
Trichuris trichiura. The library was constructed from
Trichuris muris adults (Edinburgh 'E' strain) maintained
in mice, and was provided by Dr. Richard Grencis,
University of Manchester."
BASE COUNT      102 a      80 c      99 g      62 t      10 others
ORIGIN

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:        0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-10-048-197-2 x BM277543/rev ..

Align seg 1/1 to reverse of: BM277543 from: 1 to: 353

19 ValThrGlyCysValSerThrGly 26
|||||
249 GTACACGCTGTGTTCGACTGCG 226

seq_name: gb_gss:BM033149

seq_documentation_block:
LOCUS      BH033149              358 bp      DNA      linear      GSS 17-JUL-2001
DEFINITION RPCI-24-294H20.TV RPCI-24 Mus musculus genomic clone RPCI-24-294H20
, DNA sequence.
ACCESSION  BH033149
VERSION    BH033149.1 GI:14804691
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 358)
REFERENCE   Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akiret,B., Levins,M.,
AUTHORS     Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
            Mouse BAC End Sequences from Library RPCI-24
            Unpublished (1999)
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
            page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
            Place: 294 row: H column: 20
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source             1..358
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="RPCI-24-294H20"
                        /clone_lib="RPCI-24"
                        /sex="Male"
```

```
/cell_type="Spleen/Brain"
/notes="Vector: pTRABAC1; site_1: BamHI; site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTRABAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      99 a      60 c      76 g      123 t
ORIGIN

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:        0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-10-048-197-2 x BH033149/rev ..

Align seg 1/1 to reverse of: BH033149 from: 1 to: 358

86 LeuSerLeuIleIleSerPheLeu 93
|||||
307 CTGCACTCATATATTCATTCCTT 284

seq_name: gb_gss:BH072663

seq_documentation_block:
LOCUS      BH072663              366 bp      DNA      linear      GSS 18-JUL-2001
DEFINITION RPCI-24-228C8.TV RPCI-24 Mus musculus genomic clone RPCI-24-228C8,
DNA sequence.
ACCESSION  BH072663
VERSION    BH072663.1 GI:14892260
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 366)
REFERENCE   Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akiret,B., Levins,M.,
AUTHORS     Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
            Mouse BAC End Sequences from Library RPCI-24
            Unpublished (1999)
            Other GSSs: RPCI-24-228C8.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
            page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
            Place: 228 row: C column: 8
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source             1..366
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="RPCI-24-228C8"
                        /clone_lib="RPCI-24"
                        /sex="Male"
                        /cell_type="Spleen/Brain"
                        /notes="Vector: pTRABAC1; site_1: BamHI; site_2: BamHI;
                        RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                        library was cloned in the pTRABAC1 cloning vector at the
                        BamHI sites using MboI partially digested male C57BL/6J
                        DNA."
```

```

BASE COUNT      56 a      76 c      68 g      166 t
ORIGIN
alignment_scores:
    Quality:      8.00      Length:      8
    Ratio:        1.000     Gaps:      0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x BH072663
Align seg 1/1 to: BH072663 from: 1 to: 366
86 LeuSerIeuIleIleSerPheIeu 93
|||||
166 CTGAGTTTCATATTTCTTCTC 189
seq_name: gb_est1:AM121899
seq_documentation_block:
LOCUS      AM121899              368 bp      mRNA      linear      EST 22-OCT-1999
DEFINITION UI-M-BH2.3-3-301-f-07-0-UI S1 NIH_BMAP_M.S3.3 Mus musculus cDNA clone
ACCESSION  AM121899
VERSION     AM121899.1 GI:6097273
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 368)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mestr@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized brain stems library cDNA library preparation: M.B.
            Soares Lab Clone Distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            Seg primer: M13 Forward
            POLA=yes.
FEATURES
    source
        1..368
            Location/Qualifiers
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UI-M-BH2.3-301-f-07-0-UI"
                /clone_1lb="NIH_BMAP_M.S3.3"
                /dev_stage="27-32 days"
                /lab_host="DH10B (Life Technologies)"
                /note="Vector: pTR73D-Pac (Pharmacia) with a modified
                polylinker. Site 1: Not I; Site 2: Eco RI; The
                NIH_BMAP_M.S3.3 library is a subtracted library of a
                series. ultimately derived from a mixture of individually
                tagged normalized libraries from ten regions of the mouse
                brain (cerebellum, brain stems, olfactory bulbs,
                hypothalamus, cortex, amygdala, basal ganglia, pineal
                gland, striatum, hippocampus) after a series of

```

```

subtractions to reduce the representation of cDNAs from
which ESTs had already been generated. The following
serially subtracted libraries were generated in this
process: NIH_BMAP_M.S3.3, NIH_BMAP_M.S2, NIH_BMAP_M.S1.
The subtracted library (NIH_BMAP_M.S3.3) was constructed
as follows: PCR amplified cDNA inserts from NIH_BMAP_M.S2
clones from which 3' ESTs had been derived was used as a
driver in a hybridization with the NIH_BMAP_M.S2 library
in the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the
NIH_BMAP_M.S3.3 library. This procedure has been
previously described (Bonaldo, Lennon and Soares, Genome
Research 6:791-806, 1996)
TAG_LIB=NIH_BMAP_M.S3.3
TAG_TISSUE=brain-stems
TAG_SEQ=TCATT
BASE COUNT      119 a      79 c      63 g      107 t
ORIGIN
alignment_scores:
    Quality:      8.00      Length:      8
    Ratio:        1.000     Gaps:      0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x AM121899/rev
Align seg 1/1 to reverse of: AM121899 from: 1 to: 368
22 CysValSerThrGlyAsnValAla 29
|||||
243 TGTGTCAGTACTGCAATGTGGCC 220
seq_name: gb_gss:AZ301765
seq_documentation_block:
LOCUS      AZ301765              371 bp      DNA      linear      GSS 23-AUG-2000
DEFINITION AF-F4F Resistance gene analog sequences of soybean (nonTIR-NBS-LRR
            type) similar to AF060192. Glycine max genomic, DNA sequence.
ACCESSION  AZ301765
VERSION     AZ301765.1 GI:9886539
KEYWORDS    GSS.
SOURCE      soybean.
ORGANISM    Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1 (bases 1 to 371)
AUTHORS     Penula,S. and Young,N.D.
TITLE       Resistance gene analog sequences of soybean, Glycine max
            (nonTIR-NBS-LRR type)
JOURNAL     Unpublished (2000)
COMMENT     Contact: Young, Nevin D
            Department of Plant Pathology
            University of Minnesota
            495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul MN 55108, USA
            Tel: 612 625 2225
            Fax: 612 625 9728
            Email: nevin@tc.umn.edu
            nonTIR-NBS-LRR cga sequence of soybean on linkage group F.
            Seg primer: M13F
            Class: RFLP probe.
FEATURES
    source
        1..371
            Location/Qualifiers
                /organism="Glycine max"
                /cultivar="Fairbairn"
                /db_xref="taxon:3847"
                /clone_1lb="Resistance gene analog sequences of soybean

```

(nonTR-NBS-LRR type) similar to AF060192.*
 /tissue_type="cotyledon leaves"
 /dev_stage="cotyledon stage"
 /note="Vector: PCR2.1-TOPO; Site 1: EcoRI; Soybean genomic DNA (cv. Faribault) was amplified by PCR with primers corresponding to the 'P-loop' and the region preceding the 'GTPA' motif of the nucleotide-binding site (NBS) of soybean Genbank accession AF060192. PCR products of approximately 450 bp were cloned into PCR2.1-TOPO vector, and sequenced with M13 universal primers."
 BASE COUNT 106 a 61 c 92 g 98 t 14 others
 ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AZ301765 ..
 Align seg 1/1 to: AZ301765 from: 1 to: 371
 50 GlnGluLeuSerArgpGly 57
 ||||||||||||||||||||
 224 CAAAGAGATTCTCTCGTTTGA 247

seq_name: gb_gss:AZ649818

seq_documentation_block:
 LOCUS AZ649818 376 bp DNA linear GSS 14-DEC-2000
 DEFINITION IM0519H08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0519H08 R, DNA sequence.
 ACCESSION AZ649818
 VERSION AZ649818.1 GI:11783680
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 376)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A., and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb Plasmid inserts
 Unpublished (2000)
 JOURNAL Contact: Robert B. Weiss
 COMMENT University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0519 row: H column: 08
 Seq primer: CACACAGAAACAGCATGACG
 Class: plasmid ends
 High quality sequence stop: 376.
 Location/Qualifiers
 1..376
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0519H08"
 /clone_11b="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42uv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (ql1473211419blAF129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 89 a 108 c 46 g 133 t
 ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AZ649818 ..
 Align seg 1/1 to: AZ649818 from: 1 to: 376
 86 LeuSerIleuIleSerPheIeu 93
 ||||||||||||||||||||
 131 CTCCTCTGATTAATTCCTTTT 154

seq_name: gb_est1:BB816077

seq_documentation_block:
 LOCUS BB816077 385 bp mRNA linear EST 19-NOV-2001
 DEFINITION BB816077 RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus musculus cDNA clone G730032123 3', mRNA sequence.
 ACCESSION BB816077
 VERSION BB816077.1 GI:16988706
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 385)
 Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akhiba,S., Tanaka,T., Tomaru,A., Toyota,T., Watanabe,A., Yasunishi,A., Muramatsu,M., and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
 Unpublished (2001)
 JOURNAL Contact: Yoshihide Hayashizaki
 COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new

FEATURES	High quality sequence stop: 377.
source	Location/Qualifiers
	1. 386
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:1239127"
	/clone_lib="NCI_CGAP_SSI"
	/tissue_type="synovial sarcoma"
	/lab_host="SOLR (kanamycin resistant)"
	/note="Vector: Bluescript SK-; Site.1: EcoRI; Site.2: XhoI"
	/Cloned unidirectionally. Primer: Oligo dt. Synovial
	sarcoma, 5' adaptor sequence: 5' GATTTCGGCAGCAG 3' 3'
	adaptor sequence: 5' CTCGAGCTTTTCTTTTCTTTT 3' Average
	insert size: 1.4 kb. "
BASE COUNT	73 a 62 c 72 g 179 t

BASE COUNT ORIGIN	73 a	62 c	72 g	179 t
----------------------	------	------	------	-------

US-10-048-197-2 x AA720873 ..

```

Quality:      8.00      Length:      8
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000     Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AA720873      ..

Align seg 1/1 to: AA720873 from: 1 to: 386

51 Glu11SerSerArgPheCysSer 58
|||||
317 GAGATTTCTTCTAGGTTTGATCC 340

```

```
seq_name: gb_est1:AA872067
seq_documentation_block:
docid: 1107057
```

LOCUS	392 bp	mRNA	linear	EST 25-MAR-1998
DEFINITION	AA87206/			
	ob65g905.s1	NCI CGAP GCBI	Homo sapiens	CDNA clone IMAGE:1336280 3'

```

      51 GAGTGGGCGATGCTGGTTGACATCC 58
          |||
      317 GACATTTCCTTAGGTGGATGCC 340

seq_name: gb_est1:A872067

seq_documentation_block:
LOCUS       A872067               392 bp    mRNA           linear     EST-25-MAR-19
DEFINITION  ob55g95.s1 NCI_CGAP GCBI Homo sapiens cDNA clone IMAGE:1336280 3'
             similar to gb:M55053 CYTOCHROME P450 1A2 (HUMAN);contains Alu
             repetitive element;contains element LI repetitive element ; , mRNA
             sequence.
ACCESSION   A872067

```

ACCESSION	AA872067	
VERSION	AA872067.1	GI:2968245
KEYWORDS	EST.	

LOCUS AA872067 392 bp mRNA linear EST 25-MAR-1998
DEFINITION O55905.s1 NCI CGAP-GCBI Homo sapiens CDNA clone IMAGE:1336280 3'
similar to gp:M55053 CYTOCHROME P450 1A2 (HUMAN); contains A1u
repetitive element; contains element L1 repetitive element ; mRNA
sequence.
ACCESSION AA872067
VERSION AA872067.1 GI:2968245
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 392)

Eumariyota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP)

JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be found at <http://ccgap.cancer.gov>

```

FEATURES
    source
        www-bio.linn.gov/dbcp/image/image.html
        Insert Length: 1594 Std Error: 0.00
        Seq primer: -40m13 fwd. ET from Amersham
        High quality sequence stop: 304.
        Location/Qualifiers
            1..392
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:1336280"
                /clone_11b="NCI_CGAP_GCB1"

```



```

/lab_host="DH10B"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI. 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marcil (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
15'-TGTACCAATCTGAGAGTGGAGCGCCGCTCATTTTCTTTTCTTTT-3'
1. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT      118 a      105 c      109 g      60 t
ORIGIN

alignment_scores:
    Quality:      8.00      Length:      8
    Ratio: 1.000    Gaps:      0
    Percent Similarity: 100.000    Percent Identity: 100.000

Alignment_block:
US-10-048-197-2 x AA872067/rev ..

Align seg 1/1 to reverse of: AA872067 from: 1 to: 392
      9 ThrleSerAlaLeuThrAla 16
      |||||||||||||||||||
      347 ACAATCTCTGCTGCTGCTGCTGCA 324

seq_name: gb_est2:BJ069087

seq_documentation_block:
LOCUS      BJ069087              397 bp      mRNA      linear      EST 11-DEC-2001
DEFINITION BJ069087 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL051116 5', mRNA sequence.
ACCESSION  BJ069087
VERSION     BJ069087.1 GI:17494784
KEYWORDS    EST.
SOURCE      African clawed frog.
ORGANISM    Xenopus laevis
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
            Xenopodinae; Xenopus.
            1 (bases 1 to 397)
REFERENCE   1 (bases 1 to 397)
AUTHORS    Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara
            Y.
TITLE       Expressed genes in X. laevis embryo
JOURNAL     Unpublished (2001)
COMMENT     Contact: Tadasi Shin-I
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-556-81-6856
            Fax: 81-559-81-6855
            Email: tshin@genes.nig.ac.jp.
            Location/Qualifiers
FEATURES             source
     1..397
     /organism="Xenopus laevis"
     /db_xref="taxon:8355"
     /clone_lib="NIBB Mochii normalized Xenopus tailbud
     library"
     /tissue_type="whole embryo"
     /dev_stage="stage 25"
     92 t      1 others

BASE COUNT      118 a      81 c      105 g      92 t      1 others
ORIGIN

Alignment_scores:
    Quality:      8.00      Length:      8

seq_name: gb_gss:AZ046905

seq_documentation_block:
LOCUS      AZ046905              406 bp      DNA      linear      GSS 16-MAR-2000
DEFINITION nbe00088N06f CUGI Rice BAC Library (ECORI) Oryza sativa genomic
            clone OSJNB00088N06f, DNA sequence.
ACCESSION  AZ046905
VERSION     AZ046905.1 GI:7250674
KEYWORDS    GSS.
SOURCE      Oryza sativa.
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
            1 (bases 1 to 406)
REFERENCE   1 (bases 1 to 406)
AUTHORS     Wing,R.A. and Dean,R.A.
TITLE       A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL     Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GTTAAACGACGGCCAGTG
            Class: BAC ends
            High quality sequence start: 20
            High quality sequence stop: 353.
            Location/Qualifiers
FEATURES             source
     1..406
     /organism="Oryza sativa"
     /strain="Japonica"
     /cultivar="Nipponbare"
     /db_xref="taxon:4530"
     /clone_lib="OSJNB00088N06f"
     /clone_lib="CUGI Rice BAC Library (ECORI)"
     /tissue_type="leaf"
     /lab_host="E. coli DH10B"
     /Note="Vector: pBACindigo; site_1: EcoRI; Site_2: EcoRI;
     Rice is the most important food crop in the world. Half of
     the world population, especially those inhabiting highly
     populated areas of the humid tropics and subtropics, rely
     on rice as their primary source of carbohydrate.
     Monocotyledonous rice is a diploid plant (2n=24) with a
     haploid genome equivalent of 431 Mbp (Arumuganathan and
     Earle, 1991). The relatively small genome of rice, three
     times larger than that of Arabidopsis, makes it suitable
     for genomic studies. In order to facilitate positional
     cloning, physical mapping and genome sequencing of rice,
     we have constructed a BAC library from Oryza sativa.
     Nipponbare variety using EcoRI as the cloning enzyme. The
     library contains 55,296 clones with an average insert size
     of 121 kb providing approximately 15 haploid genome
     equivalents. The deep coverage allows the isolation a
     particular sequence with a probability of 99.9 %. Three
     high density filters, each containing 18,432 clones
     (doubly spotted), represent the whole library for colony
     screening and can be requested from the Clemson University
     BAC/EST Resource Center (www.genome.clemson.edu)."
```

```

ORIGIN
alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x AZ046905 ..

Align seg 1/1 to: AZ046905 from: 1 to: 406

86 LeuSerLeu1le1eserPhaleu 93
|||||
59 TTGAGTTAATATATTCCTTTTGG 82

seq_name: gb_gss:AQ088615
seq_documentation_block:
LOCUS AQ088615 409 bp DNA linear GSS 26-AUG-1998
DEFINITION HS-2208, BL D02 MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-2208 Col-3 Row-H, DNA sequence.
ACCESSION AQ088615
VERSION AQ088615.1 GI:3457526
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 409)
Mahalas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
COMMENT Contact: Mahalas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence tagged Connector
Plate: 2208 row: H column: 3
Class: BAC ends
High quality sequence stop: 409.
FEATURES
    source
        1..409
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="CIT Approved Human Genomic Sperm Library D"
            /sex="male"
            /note="Organ: sperm; Vector: pBelOBAC11; BAC clones in
            E-Coli DH10B"
BASE COUNT 94 a 107 c 93 g 114 t 1 others
ORIGIN
alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x AQ088615 ..

Align seg 1/1 to: AQ088615 from: 1 to: 409

7 ArgLeuThrIleSerAlaLeuLeu 14

```

```

|||||
76 AGGTTACTATATCTGCTTGGCTG 99

seq_name: gb_est2:N39807
seq_documentation_block:
LOCUS N39807 410 bp mRNA linear EST 22-JAN-1996
DEFINITION yx93b05.r1 Soares melanocyte 2NBHM Homo sapiens CDNA clone
IMAGE:269265 5', mRNA sequence.
ACCESSION N39807
VERSION N39807.1 GI:1163352
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 410)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maira, M., Parsons, V.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston
R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
High quality sequence stops: 356
Source: IMAGE Consortium, LNL.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: T7
High quality sequence stop: 356.
FEATURES
    source
        1..410
            /organism="Homo sapiens"
            /db_xref="GDB:3878907"
            /db_xref="taxon:9606"
            /clone_image="269265"
            /clone_lib="Soares melanocyte 2NBHM"
            /sex="Male"
            /tissue_type="melanocyte"
            /lab_host="DH10B (ampicillin resistant)"
            /note="Vector: pT73D (Pharmacia) with a modified
            polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
            was primed with a Not I - oligo(dT) primer [5',
            TGTTACCAATCTGACAGTGGAGCGCCGAGTGTGTGTGTGTGT 3'],
            double-stranded cDNA was size selected, ligated to Eco RI
            adapters (Pharmacia), digested with Not I and cloned into
            the Not I and Eco RI sites of a modified pT73 vector
            (Pharmacia). Library constructed by Bento Soares and
            M. Fatima Bonaldo. RNA from normal foreskin melanocytes
            (FS374) was kindly provided by Dr. Anthony P. Albino."
BASE COUNT 127 a 88 c 72 g 122 t 1 others
ORIGIN
alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x N39807 ..

Align seg 1/1 to: N39807 from: 1 to: 410

11 SerAlaLeuThrAlaLeuLeu 18
|||||
91 TCAGCGCTTCTTACAGCTTGGCTG 114

```

```

seq_name: gb_est2:BU072077
seq_documentation_block:
LOCUS BU072077 411 bp mRNA linear EST 11-DEC-2001
DEFINITION BU072077 NIBB Mochi1 normalized Xenopus tailbud library Xenopus
laevis cDNA clone XL096h02 5', mRNA sequence.
ACCESSION BU072077
VERSION BU072077.1 GI:17502266
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 411)
Kitsayama,A., Terasaka,C., Mochi1,M., Ueno,N., Shin-I,T. and Kohara
Y.
REFERENCE
AUTHORS Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
FEATURES
source
1. 411
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone_1lb="NIBB Mochi1 normalized Xenopus tailbud
library"
/tissue_type="whole embryo"
/dev_stage="stage 25"
BASE COUNT 126 a 85 c 107 g 91 t 2 others
ORIGIN
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x BU072077/rev ..
Align seg 1/1 to reverse of: BU072077 from: 1 to: 411
11 SerAlaLeuLeuThrAlaLeuLeu 18
|||||
260 TCTGCTTCTTCACGACACTGCTG 237
seq_name: gb_gss:AZ132522
seq_documentation_block:
LOCUS AZ132522 427 bp DNA linear GSS 02-JUN-2000
DEFINITION OSJNB0070H11f CUGI Rice BAC Library (ECORI) Oryza sativa genomic
clone OSJNB0070H11f, DNA sequence.
ACCESSION AZ132522
VERSION AZ132522.1 GI:8211144
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eurharidodeae; Oryzae; Oryza.
1 (bases 1 to 427)
REFERENCE
AUTHORS Wing,R.A. and Dean,R.A.
JOURNAL A BAC End Sequencing Framework to Sequence the Rice Genome
COMMENT Unpublished (1998)
Contact: Wing RA
Clemson University Genomics Institute

```

```

Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: GTAAACGACGCCAGTG
Class: BAC ends
High quality sequence start: 23
High quality sequence stop: 399.
FEATURES
source
1. 427
/organism="Oryza sativa"
/strain="japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone_1lb="CUGI Rice BAC Library (ECORI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBACindigo; Site_1: EcorI; Site_2: EcorI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa,
Nipponbare variety using EcorI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT 126 a 102 c 90 g 109 t
ORIGIN
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x AZ132522 ..
Align seg 1/1 to: AZ132522 from: 1 to: 427
15 ThrAlaLeuLeuValThrGlyCys 22
|||||
334 ACTGCATTGCTGTCACGCTGCTG 357
seq_name: gb_est2:BE685620
seq_documentation_block:
LOCUS BE685620 430 bp mRNA linear EST 11-SEP-2000
DEFINITION uS56d10.x1 Soares_thymus_2NDWT Mus musculus cDNA clone
IMAGE:3375955 3', mRNA sequence.
ACCESSION BE685620
VERSION BE685620.1 GI:10073296
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 430)
REFERENCE
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

```

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other_ESTS: uu56d10.y1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:1085559.

FEATURES
source Location/Qualifiers
1..430

/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="taxon:10090"
/clone="IMAGE:3375955"
/clone_lib="Soares_thymus_2nbmt"
/sex="male"
/tissue_type="thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGACAGTGGAGCGCCGCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 153 a 90 c 77 g 110 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x BE685620/rev ..

Align seg 1/1 to reverse of: BE685620 from: 1 to: 430

22 CysValSerThrGlyAsnValAla 29
|||||
236 TGTGTGACTGCTGGAATGTGACC 213

OM of: US-10-048-197-2 to: N_Geneseq_032802:* out_format : pfs
Date: Sep 18, 2002 10:24 AM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL-frame-p2n.model -DEV-xlp
-O/cg2_1/USPRO.spool/US10048197/runatc_17092002_135957_23607/app_query.fasta_1.168
-DB-N_Geneseq_032802 -QMT-fastap -SUFFIX-oli.rng -GAPOP-4.500
-GAPOP-4.500 -MINMATCH-0.100 -LOOPCL-0.000 -LOOPEXT-0.000
-GAPOP-6.000 -FCAPEXT-7.000 -FCAPEXT-60.000 -FCAPEXT-60.000
-DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-oliigo
-TRANS-human40.cdi -LIST-45 -DOCALIGN-200 -THR_SCORE-quality
-THR_MIN-1 -ALIGN-40 -MODE-LOCAL -OUTFMT-pfs -NORM-ext
-HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000
-USPR-US10048197.@CGN1.1.523 -NCPU-6 -ICPU-3 -LONGLOG
-DEV_TIMEOUT-120 -WARN_TIMEOUT-30 -NO_XLPXY -WAIT -THREADS-1

Search information block:

Query: US-10-048-197-2

Query length: 111

Database: N_Geneseq_032802:*

Database sequences: 1736436

Search time (sec): 277.840000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF59778 +		111.00	1992.65	1.2e-102	3
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF28511 +		111.00	1951.02	2.5e-100	3
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF65480 +		8.00	137.29	26.74	2
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF89985 +		8.00	132.81	47.50	4
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA1999.DAT:AAF13630 +		8.00	132.75	47.88	4
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2002.DAT:ABK16762 +		8.00	131.63	55.26	5
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF05525 +		8.00	129.41	73.44	7
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF062008 +		8.00	128.28	84.93	8
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF53351 +		8.00	128.07	87.28	9
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2000.DAT:AAF39230 +		8.00	126.48	107.02	11
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001B.DAT:AAF81868 +		8.00	125.18	126.38	11
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF15929 +		8.00	120.14	241.23	2
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF52443 +		8.00	118.45	299.58	3
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF16230 +		8.00	116.11	404.42	4
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF16231 +		8.00	116.11	404.58	4
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF16232 +		8.00	116.07	406.33	4
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001B.DAT:ABF05316 +		8.00	111.69	712.60	8
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:ABF18935 +		8.00	110.29	852.67	11
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:ABF19362 +		8.00	110.29	852.67	11
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF136875 +		8.00	108.94	1.0e+03	1
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF19852 +		8.00	103.35	2.1e+03	2
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF36105 +		8.00	101.76	2.5e+03	3
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF31532 +		8.00	101.76	2.5e+03	3
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF31538 +		8.00	101.76	2.5e+03	3
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001B.DAT:AAF04932 +		8.00	98.10	4.1e+03	5
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF28516 +		8.00	84.15	2.4e+04	6
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF41226 +		8.00	84.15	2.4e+04	6
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF68595 +		8.00	84.15	2.4e+04	6
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF72068 +		7.00	125.31	124.35	1
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF20495 +		7.00	125.31	124.35	1
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF46634 +		7.00	125.31	124.35	1
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF52470 +		7.00	125.31	124.35	1
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF38752 +		7.00	122.78	171.99	16
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001B.DAT:AAF54847 +		7.00	122.49	202.99	1
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001B.DAT:AAF54847 +		7.00	121.49	202.99	1
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF84463 +		7.00	121.22	209.95	2
/SIDSI/gcgdata/geneseq/geneq-n-emb1/NA2001A.DAT:AAF82842 +		7.00	119.58	259.31	2

seq_name: /SIDSI/gcgdata/geneq/geneq-n-emb1/NA2001A.DAT:AAF59778
seq_documentation_block:
ID AAF59778 standard; DNA: 336 BP.
XX
AC AAF59778;
XX
DT 02-MAY-2001 (first entry)
XX
DE Moraxella catarrhalis strain ATCC43617 BASB122 DNA.
XX
KW BASB122 protein; strain ATCC43617; antigen; antibody; vaccine;
KW genetic immunisation; infection; upper respiratory tract; otitis media;
KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; antibacterial; auditory; ds.
XX
OS Moraxella catarrhalis.
XX
PN WO200109337-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000MO-EP07365.
XX
PR 30-JUL-1999; 99GB-0018034.
XX
PR 30-JUL-1999; 99GB-0018036.
XX
PA (SMIR) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonard J;
XX
DR WPT. 2001-159874/16.
XX
P-PSDB: AAB60640.
XX
New BASB122 and BASB124 polypeptides and polynucleotides from Moraxella
catarrhalis strain ATCC 43617, useful as therapeutic agents or vaccines
against bacterial infections, e.g. otitis media or pneumonia -
XX
Claim 13; Page 66; 75pp; English.
PS
The invention relates to the Moraxella catarrhalis strain ATCC43617
and to DNA encoding them (AAF59778 and AAF59779, respectively). The
invention also relates to immunogenic fragments of the BASB122 and
BASB124 proteins, expression vectors and host cells comprising BASB122
or BASB124 nucleic acids, the recombinant production of BASB122 or
BASB124, vaccine compositions comprising the BASB122 or BASB124 proteins
or nucleic acids, an antibody against BASB122 or BASB124, therapeutic
compositions comprising an anti-BASB122 or anti-BASB124 antibody, and a
method of identifying a Moraxella catarrhalis infection via the
detection of BASB122 or BASB124 proteins or antibodies. The vaccine
compositions of the invention are useful as prophylactic or therapeutic
agents against Moraxella catarrhalis infections in mammals, particularly
humans. Moraxella catarrhalis is a Gram negative bacterium frequently
isolated from the human upper respiratory tract, which is responsible for
several pathological conditions. It is responsible for about 15% of
otitis media cases in children (which can lead to temporary or permanent
hearing loss). It also causes pneumonia in elderly people, and sinusitis,
nosocomial infections and, less frequently, invasive diseases. BASB122 or
BASB124 proteins or nucleotides may additionally be used in screening for
novel antibacterial compounds, and in the diagnosis and staging of
infections. The present sequence represents DNA encoding the Moraxella
catarrhalis strain ATCC43617 BASB122 protein.
CC
Sequence 336 BP; 106 A; 65 C; 61 G; 104 T; 0 other;
XX

alignment_scores:
Quality: 111.00 Length: 111
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AAF59778 ..

Align seg 1/1 to: AAF59778 from: 1 to: 336

```
1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
|||||
1 AGTTAAATCATCATATATTCGCTGACATTCCTTCCGCTTACGAGCGCACT 50
17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGlnGln 34
|||||
51 TTGGTAAACAGGTGTGTCTTCTACTGTTAATGTTGCAATGAAAGAGCAA 100
34 sngLglnThrIleGlnThrIleIleIleLysGlyLysThrAsnLysGln 50
|||||
101 ACCAACAAACCATTTGACGACAGCATCATTAAGGCGCAAGCAATTAAGCAA 150
51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVal 67
|||||
151 GAGATTCTCTAGATTTGTTGTTCTGCTGATGACATCTCTTTATGATAGT 200
67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 84
|||||
201 GGTAAATAAATTTGACATACCGCATACCTGCTCCAAACCGATGGCAAG 250
84 lLueIleuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
|||||
251 AAATTTATCCCTTATATTTCTTTCTTGGGTAAGCAATACATACAGCA 300
101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
|||||
301 AAGAACTTGTCATTTTATTTTGACAGCAAAAGCA 333
```

seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF28551

seq_documentation_block:
ID AAF28551 standard; DNA; 94750 BP.

```
AC AAF28551;
XX
XX 04-APR-2001 (first entry)
XX
XX Genomic fragment #38.
XX
XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
XX bronchopulmonary; endocarditis; meningitis; ss.
XX
XX Moraxella catarrhalis.
XX
XX WO200078968-A2.
XX
XX 28-DEC-2000.
XX
XX 16-JUN-2000; 2000WO-US16649.
XX
XX 18-JUN-1999; 99US-0140121.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Lagace RE, Patterson C, Berg KL;
XX
XX WPI; 2001-041427/05.
XX
XX Genomic library for identifying diagnostic and therapeutic
XX PT compositions, and for identifying virulence factors, regulatory
XX PT elements and drug targets, comprises Moraxella catarrhalis nucleic
```

PT acids -
XX
XX
PS Claim 1, Page 415-436; 545pp; English.
XX
XX The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specifications e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
XX Sequence 94750 BP; 27128 A; 18811 C; 21444 G; 27367 T; 0 other.

alignment_scores:
Quality: 111.00 Length: 111
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AAF28551 ..

Align seg 1/1 to: AAF28551 from: 1 to: 94750

```
1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
|||||
75267 AGTTAAATCATCATATTCGCTGACATTCCTTCCGCTTACGAGCGCACT 75316
17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGlnGln 34
|||||
75317 TTGGTAAACAGGTGTGTCTTCTACTGTTAATGTTGCAATGAAAGAGCAA 75366
34 sngLglnThrIleGlnThrIleIleIleLysGlyLysThrAsnLysGln 50
|||||
75367 ACCAACAAACCATTTGACGACAGCATCATTAAGGCGCAAGCAATTAAGCAA 75416
51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVal 67
|||||
75417 GAGATTCTCTAGATTTGTTGTTCTGCTGATGACATCTCTTTATGATAGT 75466
67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 84
|||||
75467 GGTAAATAAATTTGACATACCGCATACCTGCTCCAAACCGATGGCAAG 75516
84 lLueIleuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
|||||
75517 AAATTTATCCCTTATATTTCTTTCTTGGGTAAGCAATACATACAGCA 75566
101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
|||||
75567 AAGAACTTGTCATTTTATTTTGACAGCAAAAGCA 75599
```

seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH65480

seq_documentation_block:
ID AAH65480 standard; DNA; 261 BP.

```
AC AAH65480;
XX
XX 26-SEP-2001 (first entry)
XX
XX C glutamicum coding sequence fragment SEQ ID NO: 515.
XX
XX Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.
XX
XX Corynebacterium glutamicum.
XX
XX EPI108790-A2.
XX
XX
```

```

XX 20-JUN-2001.
PD 18-DEC-2000; 2000EP-0127688.
XX 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX (KYOM ) KYOMA HAKKO KOGYO KK.
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.
DR P-PSDB: AAG90261.
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX Claim 8: SEQ ID NO: 515; 246bp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 261 BP; 52 A; 77 C; 54 G; 78 T; 0 other:
SO

```

alignment_scores:

Quality:	8.00	Length:	8
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

```

US-10-048-197-2 x AAH65480 ..
Align seg 1/1 to: AAH65480 from: 1 to: 261

```

```

11 SerAlaLeuLeuThrAlaLeuLeu 18
|||||
95 TCCGCTTGTCTCACAGCCCTTCTT 118

```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI89885

seq_documentation_block:

```

ID AAI89885 standard; cDNA; 479 BP.
XX
XX AAI89885;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 9945.
XX
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorders; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX

```

```

PN MO200164835-A2.
XX
XX 07-SEP-2001.
PD
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX WPI: 2001-514838/56.
DR P-PSDB: AAO09954.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
PT
XX
XX Claim 1: SEQ ID NO 9945; 1399bp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC actinin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 479 BP; 103 A; 97 C; 102 G; 176 T; 1 other:
SO

```

alignment_scores:

Quality:	8.00	Length:	8
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

```

US-10-048-197-2 x AAI89885 ..
Align seg 1/1 to: AAI89885 from: 1 to: 479

```

```

86 LeuSerLeuIleIleSerPheLeu 93
|||||
348 CTAAGCTTAATCAATTCCTTCTT 371

```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAI13630

seq_documentation_block:

```

ID AAI13630 standard; DNA; 483 BP.
XX
XX AAI13630;
XX
XX 19-MAR-1999 (first entry)
XX
XX Enterococcus faecalis genome contig SEQ ID NO: 693.
XX
XX Enterococcus faecalis; contig; detection; Enterococcal infection;
KM vaccine; attenuation; computer readable medium; ds.
XX
XX Enterococcus faecalis.
XX
XX WO9850555-A2.
XX
XX 12-NOV-1998.
XX

```

```

XX 04-MAY-1998; 98WO-US08985.
XX
XX 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Dillon RJ, Kunsch CA;
XX
XX WPI; 1999-045171/04.
XX
XX
XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
XX Claim 1; Page 1931-1932; 2084pp; English.
XX
XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosis of Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
XX Sequence 483 BP; 140 A; 84 C; 97 G; 161 T; 1 other;
SQ

```

alignment_scores:

Quality:	8.00	Length:	8
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

```

US-10-048-197-2 x AAX13630 ..

```

Align seg 1/1 to: AAX13630 from: 1 to: 483

```

7 ArgLeuThrIleSerAlaLeuLeu 14
|||||
65 CGTTTACCAATTTCAGCAGCTTTTA 88

```

seq_name: /SDSJ/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.ABK16762

seq_documentation_block:

```

ID ABK16762 standard; CDNA; 562 BP.
XX
XX ABK16762;
AC
XX
XX 14-MAR-2002 (first entry)
DT
XX
XX Human CDNA encoding G protein-coupled receptor ngPCR-Seq2664.
DE
XX
XX Human; ss; G protein-coupled receptor; ngPCR-x; thyroid disorder;
KW thyrotoxicosis; renal failure; inflammatory condition; Crohn's disease;
KW rheumatoid arthritis; autoimmune disorder; movement disorders;
KW CNS disorder; migraine; neurological disorder; anxiety;
KW manic depression; schizophrenia; epilepsy; severe mental retardation;
KW dyskinetia; Huntington's disease; Tourette's syndrome;
KW degenerative disorder; Parkinson's disease; Alzheimer's disease;
KW viral infection; HIV infection; cardiovascular disorder;
KW human immunodeficiency virus; metabolic disorder; type II diabetes;
KW obesity; anorexia; hypotension; hypertension; myocardial infarction;

```

```

KW atherosclerosis; cancer; hyperproliferative disorder; psoriasis;
KW hormonal disorder; polycystic ovarian syndrome; sexual dysfunction;
KW respiratory ailment; asthma; traumatic brain injury; angina;
KW glucose uptake disorder; metabolic disorder; angiogenesis;
KW vasculogenesis; wound healing.
XX
XX Homo sapiens.
XX
XX WO200190149-A2.
XX
XX 29-NOV-2001.
XX
XX
XX 22-MAY-2001; 2001WO-US16419.
XX
XX 22-MAY-2000; 2000US-206138P.
PR 22-MAY-2000; 2000US-206139P.
PR 02-JUN-2000; 2000US-208976P.
XX
XX (PMA ) PHARMACIA & UPJOHN CO.
XX
XX Vogell G;
XX
XX WPI; 2002-106173/14.
XX
XX P-PSDB; AAU80550.
XX
XX
XX Novel G-protein-coupled receptor polypeptides and polynucleotides
PT useful for diagnosis, prevention and treatment of viral infections,
PT cancer, metabolic, cardiovascular diseases and neurological disorders -
XX
XX Claim 4; Page 71; 138pp; English.
XX
XX The invention relates to novel G-protein-coupled receptor
XX polynucleotides, termed ngPCR-x, and the ngPCR-x proteins. The ngPCR-x
XX proteins, polynucleotides, anti-ngPCR-x antibodies, and modulators of
XX ngPCR-x are useful for the diagnosis and treatment of diseases or
XX conditions such as thyroid disorders (thyrotoxicosis), renal failure,
XX inflammatory conditions (Crohn's disease), disease related to cell
XX differentiation and homeostasis, rheumatoid arthritis,
XX autoimmune disorders, movement disorders, CNS disorders (e.g. pain
XX including migraine, stroke, psychotic and neurological disorders
XX including anxiety, mental disorders, manic depression, generalised
XX anxiety disorders, post-traumatic stress disorders, schizophrenia,
XX depression, bipolar disorder, delirium, epilepsy, dementia, severe mental
XX retardation, dyskinesias such as Huntington's disease or Tourette's
XX syndrome, attention disorders and degenerative disorders such as
XX Parkinson's, Alzheimer's disease, movement disorders including ataxias,
XX or HIV-2 (human immunodeficiency virus), metabolic and cardiovascular
XX disease and disorders (e.g. type II diabetes, impaired glucose tolerance,
XX dyslipidaemia, obesity, anorexia, hypotension, hypertension, thrombosis,
XX myocardial infarction, cardiomyopathies, atherosclerosis), proliferative
XX disease and cancers (e.g. breast, colon, lung cancer and
XX hyperproliferative disorders such as psoriasis, prostate hyperplasia),
XX hormonal disorders (male/female hormonal replacement, polycystic ovarian
XX syndrome, alopecia), sexual dysfunction, respiratory ailments such
XX as asthma, acute and/or traumatic brain injury and unstable angina due
XX to the role of T cells, inflammation, irregularities in glucose uptake
XX and metabolism, fatty acid uptake, abnormal angiogenesis, vasculogenesis,
XX wound healing. The present sequence is a CDNA encoding a ngPCR-x.
XX
XX Sequence 562 BP; 176 A; 118 C; 140 G; 128 T; 0 other;
SQ

```

alignment_scores:

Quality:	8.00	Length:	8
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

```

US-10-048-197-2 x ABK16762/rev ..

```

Align seg 1/1 to reverse of: ABK16762 from: 1 to: 562

10 lleseralaleuThralaleu 17
 |||
 66 ATTCTGCGCTCTCAGCTGCTG 43

seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH05525

seq_documentation_block:

ID AAH05525 standard; cDNA: 759 BP.

AAH05525:

26-JUN-2001 (first entry)

Human cDNA clone (5'-primer) SEQ ID NO:2360.

Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;

WPI: 2001-318749/34.

Claim 1; SEQ ID 2360; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 759 BP; 184 A; 213 C; 210 G; 149 T; 3 other:

alignment_scores: 8.00 Length: 8
 Quality: 8.00

Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AAH05525 ..

Align seg 1/1 to: AAH05525 from: 1 to: 759

12 AlaleuThralaleuVal 19

|||||

63 GCCCTGCTACAGCAGCTGCTG 86

seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA1994.DAT:AA062008

seq_documentation_block:

ID AA062008 standard; cDNA to mRNA: 885 BP.

AA062008:

26-SEP-1994 (first entry)

SSO2 gene.

Sec1 suppressor gene; homologue; secretion; biomass; yeast; ss.

Saccharomyces cerevisiae.

Key Location/Qualifiers

CDS 1..885

/*tag= a

WO9408024-A.

14-APR-1994.

06-OCR-1993; 93WO-F100402.

06-OCR-1992; 92FI-0004494.

(VALW) VALTON TEKILLINEN TUTKIMUSKESKUS.

Aalto M, Keranen S, Outola M, Penttilae M, Ronne H;

P-PSDB: AAR51274.

New sec1 suppressor gene - used for increasing prodn of secreted foreign or endogenous proteins by eukaryotic cells

Claim 3; Page 32; 51pp; English.

The yeast sec1 suppressor gene SSO2 was isolated as a suppressor of the temperature sensitive defect of the sec 1-1 mutant. When overexpressed in eukaryotic cells the gene renders the cells capable of producing increased ams. of secreted proteins, e.g. antibodies or hydrolytic enzymes. The gene may be used for more efficient biomass prodn. from raw material or efficient hydrolysis of a raw material.
 See also AA062007.

Sequence 885 BP; 320 A; 171 C; 196 G; 198 T; 0 other:

alignment_scores: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AA062008/rev ..

Align seg 1/1 to reverse of: AA062008 from: 1 to: 885

12 AlaleuLeuThraAlaleuVal 19
|||||
779 GCACCTCTTAACGCGCTGTGTGGTG 756

seq_name: /SISL1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAK53351

seq_documentation_block:

ID AAK53351 standard; cDNA; 911 BP.

AAK53351;

06-NOV-2001 (first entry)

Human polynucleotide SEQ ID NO 2880.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
tissue growth factor; immunomodulatory; cancer; leukaemia;
nervous system disorder; arthritis; inflammation; ss.

Homo sapiens.

WO200157190-A2.

09-AUG-2001.

05-FEB-2001; 2001WO-US04098.

03-FEB-2000; 2000US-0496914.

27-APR-2000; 2000US-0560875.

20-JUN-2000; 2000US-0598075.

19-JUL-2000; 2000US-0620325.

01-SEP-2000; 2000US-0654936.

15-SEP-2000; 2000US-0663561.

20-OCT-2000; 2000US-0693325.

30-NOV-2000; 2000US-0728422.

(HYSE-) HYSEQ INC.

Tang YF, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

Xue AJ, Yang Y, Wejhtman T, Goodrich R;

WPI: 2001-476283/51.

P-PSDB: AAM80218.

Nucleic acids encoding polypeptides with cytokine-like activities,

useful in diagnosis and gene therapy -

Claim 1; Page 5035; 6221pp; English.

The invention relates to polynucleotides (AAK51456-AAK53435) and the
encoded polypeptides (AAM78333-AAM80302) that exhibit activity elating to
production of other cytokines in other cell populations. The
polynucleotides and polypeptides are useful in gene therapy, vaccines or
e.g. stem cell growth factor activity, haematopoiesis regulating
activity, tissue growth factor activity, immunomodulatory activity and
activin/inhibin activity and may be useful in the diagnosis and/or
treatment of cancer, leukaemia, nervous system disorders, arthritis and
inflammation.

Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
(AAM80020) are omitted as the relevant pages from the sequence listing
were missing at the time of publication.

SO Sequence 911 BP; 205 A; 273 C; 244 G; 189 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AAK53351 ..

Align seg 1/1 to: AAK53351 from: 1 to: 911

12 AlaleuLeuThraAlaleuVal 19
|||||
178 GCCCTGCTTAACAGACACTGCTGCTGA 201

seq_name: /SISL1/gcdata/geneseq/geneseqn-emb1/NA2000.DAT:AAK39230

seq_documentation_block:

ID AAK39230 standard; DNA; 1130 BP.

AAK39230;

17-OCT-2000 (first entry)

Arabidopsis thaliana DNA fragment SEQ ID NO: 23849.

Hybridisation assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway;
metabolic pathway; promoter; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

23-APR-1999; 99US-0130891.

28-APR-1999; 99US-0131448.

30-APR-1999; 99US-0131408.

30-APR-1999; 99US-0132407.

04-MAY-1999; 99US-0132484.

05-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.

07-MAY-1999; 99US-0132487.

11-MAY-1999; 99US-0132863.

14-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

14-MAY-1999; 99US-0134370.

18-MAY-1999; 99US-0134768.

19-MAY-1999; 99US-0134941.

20-MAY-1999; 99US-0135124.

21-MAY-1999; 99US-0135353.

24-MAY-1999; 99US-0135629.

25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136392.

28-MAY-1999; 99US-0136782.

01-JUN-1999; 99US-0137222.

03-JUN-1999; 99US-0137528.

04-JUN-1999; 99US-0137502.

07-JUN-1999; 99US-0137724.

08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142820.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151530.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 8.00
Ratio: 1.000
Percent Similarity: 100.000
Length: 8
Gaps: 0
Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AAC39230 ..

Align seg 1/1 to: AAC39230 from: 1 to: 1130

13 LeuLeuThrAlaLeuLeuValThr 20
|||||
637 CTTTAAACCGCTTGGCTGCGACA 660

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAS81888

seq_documentation_block:

ID AAS81888 standard; CDNA: 1347 BP.

AC AAS81888;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #17692.

Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HXSE-) HXSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR P-PSDB; ABG17701.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 1; SEQ ID No 17692; 103pp; English.

The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of disorders or other traits to assess biodiversity
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

Sequence 1347 BP; 274 A; 331 C; 346 G; 396 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AAS81888 ..

Align seg 1/1 to: AAS81888 from: 1 to: 1347

13 LeuLeuThrAlaLeuLeuValThr 20
|||||
1100 TTGCTGACTGCATCTGTGTTACC 1123

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH15929

seq_documentation_block:

ID AAH15929 standard; CDNA: 2667 BP.

AC AAH15929;

DT 26-JUN-2001 (first entry)

DE Human CDNA sequence SEQ ID NO:14503.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI: 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length CDNA defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length CDNA

PS Claim 8; SEQ ID 14503; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602
CC full-length CDNA defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length CDNA. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length CDNA. The primers allow obtaining of the full-length
CC CDNA easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX
SQ Sequence 2667 BP; 571 A; 655 C; 760 G; 681 T; 0 other;

alignment_scores:

Quality:	8.00	Length:	8
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-10-048-197-2 x AAH15929 ..

Align seg 1/1 to: AAH15929 from: 1 to: 2667

12 AlaleuLeuThra1aleuLeuVal 19

63 GCCCTGCTAACAGCAGCTGCTGTA 86

seq_name: /SID51/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAK52367

seq_documentation_block:

ID AAK52367 standard: cDNA; 3353 BP.

XX
AC AAK52367;

XX
DT 06-NOV-2001 (first entry)

XX
DE Human polynucleotide SEQ ID NO 912.

XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

XX
OS Homo sapiens.

XX
PN WO200157190-A2.

XX
PD 09-AUG-2001.

XX
PF 05-FEB-2001; 2001WO-US04098.

XX
PR 03-FEB-2000; 2000US-0496914.

XX
PR 27-APR-2000; 2000US-0560875.

XX
PR 20-JUN-2000; 2000US-0598075.

XX
PR 19-JUL-2000; 2000US-0620325.

XX
PR 01-SEP-2000; 2000US-0654936.

XX
PR 15-SEP-2000; 2000US-0663561.

XX
PR 20-OCT-2000; 2000US-0693325.

XX
PR 30-NOV-2000; 2000US-0728422.

XX
DR WPI: 2001-476283/51.

XX
DR P-PSDB; AAM79234.

XX
PA (HYSE-) HYSEQ INC.

XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

XX
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
PS Claim 1; Page 3005-3007; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX
CC encoded polypeptides (AAM79323-AAM80302) that exhibit activity elating to
XX
CC cytokine, cell proliferation or cell differentiation or which may induce
XX
CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.

XX
SQ Sequence 3353 BP; 731 A; 802 C; 935 G; 885 T; 0 other;

alignment_scores:

Quality:	8.00	Length:	8
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-10-048-197-2 x AAK52367 ..

Align seg 1/1 to: AAK52367 from: 1 to: 3353

12 AlaleuLeuThra1aleuLeuVal 19

221 GCCCTGCTAACAGCAGCTGCTGTA 244

seq_name: /SID51/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF24443

seq_documentation_block:

ID AAF24443 standard: DNA; 4369 BP.

XX
AC AAF24443;

XX
DT 08-MAY-2001 (first entry)

XX
DE Corynebacterium thermoaminogenes plasmid pYW2.

XX
KW Plasmid; pYW2; rep; coryneform bacterium; L-amino acid; thermostable;
KW high temperature; ds.

XX
OS Corynebacterium thermoaminogenes.

XX
PN EP1076094-A2.

XX
PD 14-FEB-2001.

XX
PF 11-AUG-2000; 2000EP-0117225.

XX
PR 12-AUG-1999; 99JP-0228391.

XX
PA (AJIN) AJINOMOTO CO INC.

XX
PI Matsuzaki Y, Kimura E, Nakamatsu T, Kurahashi O, Kawahara Y;

XX
PI Sugimoto S;

XX
DR WPI: 2001-193202/20.

XX
DR P-PSDB; AAB24440.

XX
PT New plasmids derived from Corynebacterium thermoaminogenes, useful for
XX
PT improving coryneform bacteria, which can grow at elevated temperatures,
XX
PT and for producing useful substances (e.g. L-amino acids)

XX
PS Example 2; Page 21-24; 32pp; English.

XX
CC The present invention provides plasmids isolated from Corynebacterium
XX
CC thermoaminogenes encoding a rep protein. These are useful for improving
XX
CC coryneform bacteria, which are capable of growing at elevated
XX
CC temperatures and of producing substances such as L-amino acids. The
XX
CC present sequence is the coding sequence of the Coryneform bacterial
XX
CC plasmid pYW2.

SO Sequence 4369 BP; 965 A; 1252 C; 1293 G; 859 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
us-10-048-197-2 x AAF24443 ..

Align seg 1/1 to: AAF24443 from: 1 to: 4369

51 Glut1SerSerArgpHeglySer 58
|||||
2010 GAATTTCTCTCGCTCGTAGT 2033

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.ABA16230

seq_documentation_block:
ID ABA16230 standard; DNA; 4604 BP.

AC ABA16230;
XX
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 8561.
XX
KW Human; nocotropic; neuroprotective; cytosstatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; antiviral;
KW antiparkinsonian; antistickling; antianaemic; antiallergic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antidiuretic; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
XX Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 02-MAR-2000; 2000US-0184664.
PR 16-MAR-2000; 2000US-0186350.
PR 17-MAR-2000; 2000US-0190874.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217486.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234323.
PR 21-SEP-2000; 2000US-0234474.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234999.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246509.
PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX
PS Disclosure: SEQ ID NO 8561; 1701pp + Sequence Listing; English.
XX
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (AB14678-AB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 4604 BP; 1193 A; 1220 C; 1268 G; 923 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-10-048-197-2 x ABA16230 ..

Align seq 1/1 to: ABA16230 from: 1 to: 4604

17 LeuLeuValThrGlyCysValSer 24
|||||
592 CTCCTGGTCACAGGCTCGCTTTC 615

seq_name: /stidsl/gcgdata/geneseq/geneseqn_emb1/NA2001A.DAT.ABA16231
seq_documentation_block:
ID ABA16231 standard; DNA; 4606 BP.
XX
AC ABA16231;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 8562.
XX
KW Human; nootropic; neuroprotective; cyrostatic; dermatological; virucide;
KW Immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnirary;
KW antiparkinsonian; antistickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
XX Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUL-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.


```

XX AC ABL05316;
XX XX 26-MAR-2002 (first entry)
XX XX
XX XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 10430.
DE XX
XX XX Drosophila: developmental biology; cell signalling; insecticide;
KW XX pharmaceutical; gene; ss.
XX OS
XX Drosophila melanogaster.
XX PN
XX MO200171042-A2.
XX PD
XX 27-SEP-2001.
XX PF
XX 23-MAR-2001; 2001WO-US09231.
XX XX
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX XX
XX PA (PEKE ) PE CORP NY.
XX XX
XX PI Venter JC, Adams M, Li PWD, Myers EM;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABB61213.
XX XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX PS
XX Claim 1; SEQ ID NO 10430; 21pp + Sequence Listing; English.
XX XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
XX CC sequences (ABU01840-ABU16175) and the encoded proteins
XX CC (ABU57737-ABU72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pcl_sequences.
XX CC
XX SQ Sequence 8377 BP; 2677 A; 1829 C; 1525 G; 2346 T; 0 other;

alignment_scores:
    quality: 8.00      length: 8
    ratio: 1.000      gaps: 0
    percent similarity: 100.000    percent identity: 100.000

alignment_block:
US-10-048-197-2 x ABL05316/rev ..

Align seg 1/1 to reverse of: ABL05316 from: 1 to: 8377

34 ASNGINGIntHrILeGIuGIIntHr 41
|||||
2698 AATCAACAACATCGAACAAACA 2675

seq_name: /SIDSL/9csgdata/geneseq/geneseqn-emb1/NA2001A.DAT:ABA18955
seq_documentation_block:
ID ABA18955 standard; DNA; 10126 BP.
XX AC ABA18955;
XX XX
XX XX 23-JAN-2002 (first entry)
XX XX
XX XX Human nervous system related polynucleotide SEQ ID NO 11296
DE XX

```

XX Human; nootropic; neuroprotective; cytosstatic; dermatological; viroicide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnecary;
KW antiparkinsonian; antistickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-Aug-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226886.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0255719.
PR 06-DEC-2000; 2000US-0251479.

KW	neurological disease; infection; nephrotropic; gene therapy; vaccine; ds
XX	
OS	Homo sapiens.
PN	MO200159063-A2.
XX	
PD	16-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01334.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184654.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0186874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226868.
PR	22-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227009.
PR	30-AUG-2000; 2000US-0228924.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.
PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.
PR	05-SEP-2000; 2000US-0229509.
PR	05-SEP-2000; 2000US-0229513.
PR	06-SEP-2000; 2000US-0230437.
PR	06-SEP-2000; 2000US-0230438.
PR	08-SEP-2000; 2000US-0232081.
PR	12-SEP-2000; 2000US-0231958.
PR	14-SEP-2000; 2000US-0232397.
PR	14-SEP-2000; 2000US-0232398.
PR	14-SEP-2000; 2000US-0232399.
PR	14-SEP-2000; 2000US-0232400.
PR	14-SEP-2000; 2000US-0232401.
PR	14-SEP-2000; 2000US-0233063.
PR	14-SEP-2000; 2000US-0233064.
PR	21-SEP-2000; 2000US-0234223.
PR	21-SEP-2000; 2000US-0234274.
PR	25-SEP-2000; 2000US-0234977.

PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0238935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 05-DEC-2000; 2000US-0251986.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
PS
PS Disclosure; SEQ ID NO 11693; 1701bp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 10126 BP; 2749 A; 2796 C; 2248 G; 2333 T; 0 other;

alignment_scores: Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x ABA19362/rev ..

Align seg 1/1 to reverse of: ABA19362 from: 1 to: 10126

49 LysGlnGluIleSerGArgphe 56
|||||
1698 AACCAAGAAATATCAAGTACGTTT 1675

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAL36875

seq_documentation_block:
ID AAL36875 standard; DNA; 12167 BP.
XX
XX AAL36875;
AC
XX
XX 08-JAN-2002 (first entry)
DT
XX
XX Human musculoskeletal system related polynucleotide SEQ ID NO 3340.
DE
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200155367-A1.
PN
XX
XX 02-AUG-2001.
PD
XX

PF 17-JAN-2001; 2001WO-US01338.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225275.
PR 14-AUG-2000; 2000US-0225275.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226861.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235634.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239395.
PR 13-OCT-2000; 2000US-0239397.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251779.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
XX
PA
XX
PI Rosen CA, Barash SC, Ruden SM;
XX WPI: 2001-451937/48.
DR
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing

PT disorders related to the musculoskeletal system including
PT musculoskeletal cancers and also for testing and detection e.g.
PT diagnosis -
XX
XX Example 2: SEQ ID NO 3240; 781pp + Sequence Listing: English.
PS
XX The invention relates to novel genes (AAL34659-AAL37666) and proteins
CC (AB03087-AB04109) associated with the musculoskeletal system useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids, proteins,
CC antibodies and (ant)agonists are useful in the diagnosis, treatment
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune hemolytic
CC anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 12167 BP; 3252 A; 2986 C; 2599 G; 3330 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AAL36875/rev ..

Align seg 1/1 to reverse of: AAL36875 from: 1 to: 12167

88 LeuLeIleSeRPhLeuTRpval 95
|||||
10917 CTGATAATTCTTCTTCTTGGGTA 10894

seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:ABA19852

seq_documentation_block:
ID ABA19852 standard; DNA: 25955 BP.
XX
AC ABA19852;
XX
XX
DT 23-JAN-2002 (first entry)
XX
XX Human nervous system related polynucleotide SEQ ID NO 12183.
DE
XX
XX Human: nocrotropic; neuroprotective; cytoskeletal; dermatological; virucide;
KM immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KM antipariteconian; antisticking; antianaemic; antiarthritic; cancer;
KM antitumoural; hepatotropic; cerebroprotective; antiinflammatory;
KM antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KM antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KM neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
XX
PN WO200159063-A2.
XX
XX
PD 16-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229387.
PR 01-SEP-2000; 2000US-0229393.
PR 01-SEP-2000; 2000US-0229394.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246509.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251472.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
XX useful for preventing, diagnosing and/or treating nervous system
XX cancers and metastases -
XX
XX Disclosure; SEQ ID NO 12183; 1701pp + Sequence Listing; English.
XX

CC The invention relates to novel genes (ABAI1004-ABA21534) and proteins
CC (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (anti)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, Rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX SQ Sequence 25955 BP; 8559 A; 4165 C; 4808 G; 8423 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x ABAI9852 ..
Align seg 1/1 to: ABAI9852 from: 1 to: 25955

12 AIAleuLeuThAlaLeuLeuVal 19
|||||
16733 GCATTGCTCAGCTGTTGTGTGTA 16756

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA36105

seq_documentation_block:
ID AA36105 standard; DNA; 32195 BP.

XX
XX AC AA36105;
XX
XX DT 17-DEC-2001 (first entry)
XX
XX DE Human cardiovascular system antigen genomic DNA SEQ ID NO 1605.
XX

KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotrophic; dog;
KW antirheumatic; antiproliferative; cytosolic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnerrary; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; tissue regeneration;
KW anti-interfility.
XX

OS Homo sapiens.
XX

PN WO200155321-A2.
XX

PD 02-AUG-2001.
XX

XX 17-JAN-2001; 2001WO-US01340.
XX

PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225265.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231480.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-451930/48.

New Cardiovascular system related polynucleotides and polypeptides,
useful for diagnosing, treating and/or preventing disorders of the
cardiovascular system -
XX
PS Claim 1; SEQ ID NO 1605; 674pp; English.
XX
CC Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC the Cardiovascular system antigen polypeptides of the invention.
CC Cardiovascular system antigens and their associated polynucleotides are

CC useful in the diagnosis, treatment and prevention of various types of
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. A pathological condition can be determined by
CC detecting the presence or absence of a mutation in a cardiovascular
CC system antigen polynucleotide. The treatable disorders include autoimmune
CC diseases such as rheumatoid arthritis, hyperproliferative disorders such
CC as neoplasms of the breast or liver, cardiovascular disorders such as
CC cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
CC nervous system disorders such as Alzheimer's disease, infections caused
CC by bacteria, viruses and fungi, ocular disorders such as corneal
CC infection, endocrine disorders such as premature labour and infertility,
CC gastrointestinal disorders such as Crohn's disease, renal disorders such
CC as glomerulonephritis and respiratory disorders such as asthma and
CC pleurisy. The polypeptides can also be used to aid wound healing, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, to regenerate tissues and in chemotaxis.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AAS36105/rev ..

Align seg 1/1 to reverse of: AAS36105 from: 1 to: 32195

85 IleleuSerleullelleSerPhe 92

|||||
25174 ATACTATCTTAAATAATATCAATT 25151

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAS31532

seq_documentation_block:

ID AAS31532 standard; DNA: 32195 BP.

AC AAS31532;

DT 04-DEC-2001 (first entry)

DE Human DNA for a novel extracellular matrix protein, Seq ID No 611.

KW Human; secreted extracellular matrix protein; ds; immunomodulatory;
KW Anti-HIV; antianemic; antipneumatic; antisclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;
KW antilzheimer's; immune/autimmune disease; HIV infection; anaemia;
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW cancers; hyperproliferative disorder; breast neoplasm; melanoma;
KW Sezary syndrome; Gaucher's disease; neurological diseases;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW cardiac arrest; tachycardia; angina; infection; corneal infections;
KW wound healing; immunogen; gene therapy; antisense; food additive.

OS Homo sapiens.

XX WO200153368-A1.

XX 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01348.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232060.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 23-SEP-2000; 2000US-0234997.
PR 23-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX
XX MPI; 2001-465572/50.
XX
XX
XX Nucleic acid molecules encoding human secreted extracellular matrix
XX
XX
XX proteins, used in preventing, treating or ameliorating a disorder, e.g.
XX
XX Alzheimer's and Parkinson's diseases and cancers .
XX
XX
XX Claim 1; SEQ ID NO 611; 577pp; English.
XX
XX
XX The invention relates to isolated nucleic acid molecules encoding
XX
XX novel human secreted extracellular matrix proteins (SPs). The
XX
XX polynucleotides and proteins are used to prevent, treat a medical
XX
XX condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
XX
XX chickens or sheep. For example, disorders associated with decreased

CC expression of SPs. The SP polynucleotide or a vector expressing them may
CC be administered to treat diseases by gene therapy. Antisense molecules
CC may be administered to down regulate expression of SPs by binding with
CC the cells own genes and preventing their expression. The polynucleotides
CC may also be used as DNA probes in diagnostic assays. The SPs may also be
CC used as antigens to produce antibodies and to identify modulators
CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and
CC antagonists may also be used to down regulate expression and activity of
CC SP and as diagnostic agents for detecting the presence of SPs in samples.
CC The disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac
CC arrest, tachycardia and angina), infections caused by bacteria, viruses
CC and fungi and ocular disorders (e.g. corneal infections). Other uses
CC include wound healing, maintenance of organs before transplantation,
CC support of cell culture of primary tissues, modulation of for example
CC
alignment_scores: Quality: 8
 Ratio: 1.000
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x AAS31532/rev ..
Align seg 1/1 to reverse of: AAS31532 from: 1 to: 32195
55 ILeLeuSerLeuIleIleSerPhe 92
|||||
25174 ATACTACTCTTAAATATATATCAATT 25151
seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAS31538
seq_documentation_block:
ID AAS31538 standard; DNA: 32195 BP.
XX
XX AAS31538;
XX
XX
XX 04-DEC-2001 (first entry)
XX
XX
DE Human DNA for a novel extracellular matrix protein, Seq ID NO 617.
XX
XX
XX Human; secreted extracellular matrix protein; ds; immunomodulatory;
XX
XX Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiant; vascular;
XX
XX cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;
XX
XX antialzheimers; immune/autoimmune disease; HIV infection; anaemia;
XX
XX human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
XX
XX cancers; hyperproliferative disorder; breast neoplasm; melanoma;
XX
XX Sezary syndrome; Gaucher's disease; neurological diseases;
XX
XX Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
XX
XX cardiac arrest; tachycardia; angina; infection; corneal infections;
XX
XX wound healing; immunogen; gene therapy; antisense; food additive.
OS Homo sapiens.
XX
XX
XX WO200155368-A1.
XX
XX
XX 02-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US01348.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217486.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI
XX
DR
XX
XX
PT
PT
PT
XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
The invention relates to isolated nucleic acid molecules encoding novel human secreted extracellular matrix proteins (SPs). The polynucleotides and proteins are used to prevent, treat a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. For example, disorders associated with decreased expression of SPs. The SP polynucleotide or a vector expressing them may

CC be administered to treat diseases by gene therapy. Antisense molecules
CC may be administered to down regulate expression of Sps by binding with
CC the cells own genes and preventing their expression. The polynucleotides
CC may also be used as DNA probes in diagnostic assays. The Sps may also be
CC used as antigens to produce antibodies and to identify modulators
CC (agonists and antagonists) of the Sps. The anti-(Sps) antibodies and
CC antagonists may also be used to down regulate expression and activity of
CC Sps and as diagnostic agents for detecting the presence of Sps in samples.
CC The disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac
CC arrest, tachycardia and angina), infections caused by bacteria, viruses
CC and fungi and ocular disorders (e.g. corneal infections). Other uses
CC include wound healing, maintenance of organs before transplantation,
CC support of cell culture of primary tissues, modulation of for example

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AASJ1538/rev ..

Align seq 1/1 to reverse of: AASJ1538 from: 1 to: 32195

85 IlleusSerleuIleIleSerPhe 92
|||||
25174 ATACTATCTTTAATATATCTTT 25151

seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABLO4952

seq_documentation_block:

ID ABL04952 standard; CDNA: 52872 BP.

AC ABL04952;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide seq ID NO 9338.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EM;

DR WPI: 2001-656860/75.

DR P-PSDB: ABB60849.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Claim 1: SEQ ID NO 9338; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU1840-ABU16175) and the encoded proteins
CC (AAB5737-ABR7072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 52872 BP; 15532 A; 11487 C; 10803 G; 15050 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x ABL04952 ..

Align seq 1/1 to: ABL04952 from: 1 to: 52872

12 AlaleuThrAlaleuVal 19
|||||
9772 GCATTATTAAGCTTCTTCTT 9795

seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF28546

seq_documentation_block:

ID AAF28546 standard; DNA: 63563 BP.

AC AAF28546;

DT 04-APR-2001 (first entry)

DE Genomic fragment #33.

KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;

KW bronchopulmonary; endocarditis; meningitis; ss.

OS Moraxella catarrhalis.

PN WO200078968-A2.

PD 28-DEC-2000.

PF 16-JUN-2000; 2000WO-US16649.

PR 18-JUN-1999; 99US-0140121.

PA (INCY-) INCYTE GENOMICS INC.

PI Lagace RE, Patterson C, Berg KL;

DR WPI: 2001-041427/05.

PT Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids -

PS Claim 1: Page 309-324; 545pp; English.

CC The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis

CC and meningitis.
XX
SQ Sequence 63563 BP; 18440 A; 12630 C; 11985 G; 20508 T; 0 other;
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x AAF28546 ..
Align seg 1/1 to: AAF28546 from: 1 to: 63563
84 GIUlleuserleutleieser 91
|||||
45213 GAAATACATCTCTATCTCA 45236
seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH41226
seq_documentation_block:
ID AAH41226 standard; DNA: 349980 BP.
XX
AC AAH41226;
XX
DT 29-OCT-2001 (first entry)
XX
DE Pyrococcus abyssi genomic fragment #5.
XX
KM Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX
OS Pyrococcus abyssi.
XX
FH Key Location/Qualifiers
FT misc_feature 1..49980 /tag= a
FT /note= "This sequence overlaps with the 3' end of
FT misc_feature 300001..349980 /tag= b
FT /note= "This sequence overlaps with the 5' end of
FT AAH41227"
XX
PN FR2792651-A1.
XX
PD 27-OCT-2000.
XX
PF 21-APR-1999; 99FR-0005034.
XX
PR 21-APR-1999; 99FR-0005034.
XX
PA (CNRS) CNRS CENT NAT RECH SCT.
XX (IFRE) IFREMER INST FR RECH EXPL MER.
XX
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX Querellon J, Weissenbach J, Saurin W, Hellig R;
XX WPI; 2001-126236/14.
XX
PT New nucleotide sequences isolated from Pyrococcus abyssi encode
XX proteins useful in industry .
XX
PS Claim 1; Page 511-606; 1657pp; French.
XX
CC The present invention relates to the genomic sequence of Pyrococcus
XX abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
XX hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
XX vents. The present sequence is a fragment of the genomic sequence of P
XX abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41225
XX and the 3' end of this sequence overlaps with the 5' end of AAH41227. The
XX proteins of the present invention have various potential industrial uses,
XX since the proteins are stable at very high temperatures, some up to 110

CC degrees centigrade.
CC Note: This patent is in the same patent family as WO200065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAH75903-AAH75920 and AAG66436.
XX
SQ Sequence 349980 BP; 98084 A; 80447 C; 77665 G; 93784 T; 0 other;
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-10-048-197-2 x AAH41226/rev ..
Align seg 1/1 to reverse of: AAH41226 from: 1 to: 349980
9 Thrileseraleutthrala 16
|||||
104650 ACAATATCAGCCCTTTAAGCGCC 104627
seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH68525
seq_documentation_block:
ID AAH68525 standard; DNA: 349980 BP.
XX
AC AAH68525;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 7060.
XX
KM Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EPI108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX
PA (RYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI; 2001-376931/40.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
PS Disclosure; SEQ ID NO: 7060; 246pp + Sequence listing; English.
XX
CC The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from Coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a nucleic acid described
XX in the exemplification of the invention.
XX Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 349980 BP; 80289 A; 91081 C; 97378 G; 81232 T; 0 other;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAH68525/rev ..

Align seg 1/1 to reverse of: AAH68525 from: 1 to: 349980

11 SerAlaLeuThrAlaLeu 18
 |||||
 181297 TCCGCTTGCTCAGCCCTCTT 181274

seq_name: /SISL/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:ABA72068

seq_documentation_block:
 ID ABA72068 standard; DNA; 115 BP.

AC ABA72068;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #20373.

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX MO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver -

XX Claim 4; SEQ ID NO 20373; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human foetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX foetal liver. The present sequence is a single exon nucleic acid

XX probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 115 BP; 26 A; 35 C; 14 G; 40 T; 0 other;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x ABA72068 ..

Align seg 1/1 to: ABA72068 from: 1 to: 115

87 SerLeuIleIleSerPheLeu 93
 |||||
 38 TCTTGCATTATTTCCTTCTC 58

seq_name: /SISL/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAK20495

seq_documentation_block:
 ID AAK20495 standard; DNA; 115 BP.

AC AAK20495;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe SEQ ID NO: 20486.

KW Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

KW epilepsy; cancer; ss.

XX Homo sapiens.

XX MO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX brains -

XX Example 4; SEQ ID NO: 20486; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX brain. They can be used to measure gene expression in brain cell samples,

XX which may enable the diagnosis and improved treatment of nervous system

XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX epilepsy and cancers. The present sequence is one of the probes of the

XX invention.
 CC
 SQ Sequence 115 BP; 26 A; 35 C; 14 G; 40 T; 0 other;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AAK20495 ..

Align seg 1/1 to: AAK20495 from: 1 to: 115

87 Serleu1le1leSerPhelu 93
|||||
38 TCTTGTATATTCCTTCTC 58

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAK46634

seq_documentation_block:

ID AAK46634 standard; DNA; 115 BP.

AC AAK46634;

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed single exon probe SEQ ID NO: 21191.

KW Human: bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

OS Homo sapiens.

PN WO200157276-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

DR Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human bone marrow -

PS Example 4; SEQ ID NO: 21191; 658bp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in bone marrow

CC samples, which can be used to measure gene expression and treatment of cancers

CC such as lymphoma, leukemia and myeloma. The present sequence is one of

CC the probes of the invention.

XX Sequence 115 BP; 26 A; 35 C; 14 G; 40 T; 0 other;

alignment_scores: Quality: 7.00 Length: 7

Percent Similarity: 100.000 Ratio: 1.000 Gaps: 0

Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AAK46634 ..

Align seg 1/1 to: AAK46634 from: 1 to: 115

87 Serleu1le1leSerPhelu 93

|||||

38 TCTTGTATATTCCTTCTC 58

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI52470

seq_documentation_block:

ID AAI52470 standard; DNA; 115 BP.

AC AAI52470;

DT 17-OCT-2001 (first entry)

DE Probe #21156 used to measure gene expression in human placenta sample.

KW Probe; microarray; human; placenta; antenatal diagnosis;

OS genetic disorder; ss.

PN Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

DR Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -

PS Claim 25; SEQ ID NO 21156; 654bp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).

CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders.

XX Sequence 115 BP; 26 A; 35 C; 14 G; 40 T; 0 other;

alignment_scores: Quality: 7.00 Length: 7

Percent Similarity: 100.000 Ratio: 1.000 Gaps: 0

Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AAI52470 ..

Align seg 1/1 to: AAI52470 from: 1 to: 115

87 Serleu1le1leSerPhelu 93

|||||

38 TCTTGTATATTCCTTCTC 58

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAI52470

seq_documentation_block:

ID AAC98752 standard; cDNA; 162 BP.

AC AAC98752;

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen nucleotide sequence SEQ ID NO:162.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW Identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW Immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotoxic; antineoplastic; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; se.
XX
XX Homo sapiens.
OS
XX
XX WO200055351-A1.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 08-MAR-2000; 2000WO-US05883.
PE
XX
XX 12-MAR-1999; 99US-0124270.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
P1
XX
XX WPI; 2000-587534/55.
DR
XX
XX P-PSDB; AAB53995.
DR
XX
XX Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
PS
XX
XX Claim 1; Page 1323; 2104pp; English.
PS
XX
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerary, nephrotoxic, antineoplastic and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAC954007 represent sequences used in the exemplification of the present
CC invention.
CC
XX
XX Sequence 162 BP; 34 A; 50 C; 33 G; 37 T; 8 other;
SQ

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AAC98752 ..
Align seg 1/1 to: AAC98752 from: 1 to: 162
13 LeuLeuThrAlaLeuLeuVal 19
|||||
56 CTGCTNACAGCCCTTACTTGTGA 76

seq_name: /SIDSI/gcsgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAS48479
seq_documentation_block:
ID AAS48479 standard; DNA; 179 BP.
XX
XX AAS48479;

XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Klebsiella pneumoniae cellular proliferation inhibitory sequence #3.
DE
XX
XX Antisense; ss; prokaryotic cellular proliferation;
KW antibiotic; antibacterial; drug design.
XX
XX Klebsiella pneumoniae.
OS
XX
XX WO200170955-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 21-MAR-2001; 2001WO-US09180.
PE
XX
XX 21-MAR-2000; 2000US-191078P.
PR
XX
XX 23-MAY-2000; 2000US-206848P.
PR
XX
XX 26-MAY-2000; 2000US-207727P.
PR
XX
XX 23-OCT-2000; 2000US-242578P.
PR
XX
XX 27-NOV-2000; 2000US-253625P.
PR
XX
XX 22-DEC-2000; 2000US-257931P.
PR
XX
XX 16-FEB-2001; 2001US-269308P.
PA
XX
XX (ELIT-) ELITRA PHARM INC.
PI
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
PI
XX
XX WPI; 2001-611495/70.
DR
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
PT
XX
XX Claim 1; Seq ID No 1056; 511pp; English.
PS
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence is an antisense
CC oligonucleotide of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pcl_sequences.
CC
XX
XX Sequence 179 BP; 68 A; 38 C; 36 G; 37 T; 0 other;
SQ

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AAS48479 ..
Align seg 1/1 to: AAS48479 from: 1 to: 179
6 IleArgLeuThrIleSerAla 12
|||||
2 ATCAGGCTGACATCAGCGCC 22

seq_name: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT:AA548487

seq_documentation_block:

ID AA548487 standard; DNA; 193 BP.

AC AA548487;

DT 13-FEB-2002 (first entry)

DE Klebsiella pneumoniae cellular proliferation inhibitory sequence #11.

KW Antisense; ss; prokaryotic cellular proliferation;

KM antibiotic; antibacterial; drug design.

OS Klebsiella pneumoniae.

PN WO200170955-A2.

XX 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

XX WPI; 2001-611495/70.

Align seg 1/1 to: AA548487 from: 1 to: 193

6 IleArgLeuThrIleSerAla 12

|||||

55 ATCAGGCTGACAAATCAGCCGC 75

seq_name: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH84463

seq_documentation_block:

ID AAH84463 standard; DNA; 200 BP.

AC AAH84463;

DT 26-SEP-2001 (first entry)

DE E. coli growth and proliferation related DNA sequence SEQ ID NO:91.

KW Escherichia coli; growth; proliferation; microbial; antimicrobial;

KM bacterial infection; microorganism; ds.

OS Escherichia coli.

PN WO200134810-A2.

XX 17-MAY-2001.

PF 09-NOV-2000; 2000WO-US30950.

PR 09-NOV-1999; 99US-0164415.

XX (ELIT-) ELITRA PHARM INC.

PI Forsyth RA, Ohlsen K, Zyskind J;

PI WPI; 2001-335933/35.

XX Novel nucleic acids that inhibit Escherichia coli proliferation, useful

PT for screening for homologous genes and for designing expression vectors

XX Claim 1; Page 123; 522pp; English.

XX AAH84373 to AAH84499 represent Escherichia coli growth and proliferation

CC related DNA sequences (1). AAH84500 to AAH84670 encode the E. coli

CC growth and proliferation related proteins given in AG99078 and AG99830

CC to AG99999. (1) can be used as potential targets for the generation of

CC new antimicrobial agents, and for identification of compounds which

CC interact with the gene products of (1). In addition the expression of

CC (1) and the purification of the proteins, the purified proteins can be

CC used to generate reagents and screen small molecule libraries or other

CC candidate compound libraries for compounds that can be further developed

CC to yield novel antimicrobial compounds. In addition, nucleic acid probes

CC complementary to (1) that are specific for particular species of

CC microorganisms can be used to identify particular microorganism species

CC in clinical specimens, therefore, providing a rapid and dependable

CC method by which to identify the causative agents of a bacterial

CC infection. Also, antibodies generated against proteins translated from

CC mRNA transcribed from proliferation-regulated sequences can also be used

CC to screen for specific microorganisms that produce such proteins in a

CC species-specific manner. AAH84371 and AAH84670 represent sequencing

CC primers used in the isolation of E. coli growth and proliferation

CC related sequence, which are used in an example from the present

XX Sequence 200 BP; 52 A; 50 C; 49 G; 49 T; 0 other;

alignment_scores:

Quality: 7.00

Ratio: 1.000

Percent Similarity: 100.000

Length: 7

Gaps: 0

Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AA548487

alignment_block:

US-10-048-197-2 x AAH84463/rev ..

Align seg 1/1 to reverse of: AAH84463 from: 1 to: 200

11 SerAlaLeuLeuThrAlaLeu 17
|||||
56 TCCGCTCTCCTAACTGCCTTA 36

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH82842

seq_documentation_block:

ID AAC82842 standard; cDNA; 250 BP.

AC AAC82842;

DT 20-MAR-2001 (first entry)

DE Murine LMTF clone 701252210H1 cDNA.

XX LMTF: lipid metabolism transcription factor; cytosolic; hepatotropic;
KM antlerteriosclerotic; nephrotropic; cell proliferation disorder; cancer;
KM atherosclerosis; cirrhosis; hepatitis; lipid disorder; murine; ss.

OS Mus musculus.

PN WO200073443-A1.

PD 07-DEC-2000.

PF 15-MAY-2000; 2000WO-US13393.

PR 26-MAY-1999; 99US-0318978.

PA (INCY-) INCYTE GENOMICS INC.

PI Yue H, Kaser MR, Baughn MR;

DR WPI: 2001-061530/07.

PT New polynucleotide encoding lipid metabolism transcription factor is
XX useful as a probe for detecting mammalian nucleic acids in a sample

PS Claim 3: Page 46; 52pp; English.

CC This invention describes a novel purified mammalian nucleotide sequence
CC (I) encoding a lipid metabolism transcription factor (LMTF) which has
CC cytosolic, hepatotropic, antlerteriosclerotic and nephrotropic activity.
CC The probe hybridizing (I) is useful for detecting a mammalian nucleic
CC acid sequence in a sample. (I) is useful to screen a library of molecules
CC to identify at least a molecule specific to (I), comprising combining (I)
CC with a library of molecules under conditions of binding and detecting the
CC specific binding. LMTF is useful for treating or preventing a condition
CC associated with altered expression or activity of the protein, which
CC includes cell proliferation disorders such as atherosclerosis, cirrhosis,
CC hepatitis, cancer, and various lipid disorders.

SO Sequence 250 BP; 56 A; 67 C; 48 G; 78 T; 1 other;

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x AAC82842 ..

Align seg 1/1 to: AAC82842 from: 1 to: 250

15 ThrAlaLeuLeuValThrGly 21
|||||
85 ACAGCACTGTGCTAACCGGT 105

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:ABA16416

seq_documentation_block:
ID ABA16416 standard; DNA; 269 BP.

AC ABA16416;

DT 23-JAN-2002 (first entry)

DE Human nervous system related polynucleotide SEQ ID NO 8747.

XX Human: nocotropic; neuroprotective; cytosolic; dermatological; virucide;
KM immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KM antiparkinsonian; antisticking; antianaemic; antiarthritic; cancer;
KM antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
KM antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KM antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KM neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

OS Homo sapiens.

PN WO200159063-A2.

PD 16-AUG-2001.

PF 17-JAN-2001; 2001WO-US01334.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226811.

PR 22-AUG-2000; 2000US-0226868.

PR 23-AUG-2000; 2000US-0227182.

PR 30-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.

PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI, 2001-541565/60.

XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
XX useful for preventing, diagnosing and/or treating nervous system
XX cancers and metastases -

XX Disclousure: SEQ ID NO 8747; 1701pp + Sequence Listing; English.

XX The invention relates to novel genes (ABAI1004-ABAI21534) and proteins
XX (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 269 BP; 22 A; 104 C; 23 G; 120 T; 0 other;

XX Alignment scores:

XX Quality: 7.00 Length: 7
XX Ratio: 1.000 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 100.000

XX alignment_block:

XX US-10-048-197-2 x ABAI6416/rev ..

XX Align seg 1/1 to reverse of: ABAI6416 from: 1 to: 269

XX 87 Serleulleleserphelu 93

XX |||||||||||||||||||

XX 31 TCCCTCATCATCAGTTTCCTC 11

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:ABAI6419

seq_documentation_block:

XX ID ABAI6419 standard; DNA; 269 BP.

XX AC ABAI6419;

XX XX

DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 8750.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antitickling; antianaemic; antiarthritic; cancer;
KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antidiabetic; antidiabetic; antilucer; anticonvulsant; antitungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PM WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0233397.

PR 14-SEP-2000; 2000US-0233398.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0233422.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.
 XX

PI Rosen CA, Barash SC, Ruben SM;
 XX

DR WPI: 2001-541565/60.
 XX

PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX

PS Disclosure: SEQ ID NO 8750; 1701pp + Sequence Listing; English.
 XX

CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX

SO Sequence 269 BP; 22 A; 104 C; 23 G; 120 T; 0 other;

alignment_scores:

Quality:	7.00	Length:	7
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-10-048-197-2 x ABA16419/rev ..

Align seg 1/1 to reverse of: ABA16419 from: 1 to: 269

87 SerleullelleSerPhelen 93
 ||||||||||||||||||
 31 TCCCTCATCATCAGTTCTC 11

OM of: US-10-048-197-2 to: Issued_Patents_NA:* out_format : pfs
Date: Sep 18, 2002 10:19 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=framet_p2n.model -DEV=xlp  
-O=/cgn2_1/USPTO_pool/US10048197/runatc.17092002.135957.23562/app_query.fasta.1.168  
-DB=Issued_Patents_NA -QFMT=fastp -SUFFIX=011.rml -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000  
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=quality  
-THR_MIN=1 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs -NOR=ext  
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10048197 -ECGN1.1.74 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPX -WAIT -THREADS=1
```

Search information block:

```
Query: US-10-048-197-2  
Query length: 111  
Database: Issued_Patents_NA:*  
Database sequences: 38353  
Search time (sec): 77.900000
```

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation	..	885	1	5
/cgn2_6/prodata/2/1na/5A.COMB.seq:US-08-411-706-3		8.00	126.20	24.29	885	1	5		
/cgn2_6/prodata/2/1na/5B.COMB.seq:US-09-318-978-15		7.00	117.67	72.50	250	1	5		
/cgn2_6/prodata/2/1na/6A.COMB.seq:US-08-747-221B-1		7.00	114.28	112.01	401	1	5		
/cgn2_6/prodata/2/1na/6B.COMB.seq:US-08-747-221B-3		7.00	114.28	112.01	401	1	5		
/cgn2_6/prodata/2/1na/6A.COMB.seq:US-09-005-051-1		7.00	114.28	112.01	401	1	5		
/cgn2_6/prodata/2/1na/5A.COMB.seq:US-09-005-051-3		7.00	114.28	112.01	401	1	5		
/cgn2_6/prodata/2/1na/5B.COMB.seq:US-08-367-198A-4		7.00	112.02	149.58	549	1	5		
/cgn2_6/prodata/2/1na/5A.COMB.seq:US-08-367-198A-1		7.00	107.95	252.39	969	1	5		
/cgn2_6/prodata/2/1na/5B.COMB.seq:US-08-367-198A-1		7.00	107.94	252.63	970	1	5		
/cgn2_6/prodata/2/1na/6A.COMB.seq:US-08-961-083-7		7.00	107.83	256.23	985	1	5		
/cgn2_6/prodata/2/1na/5A.COMB.seq:US-08-961-083-9		7.00	105.66	338.34	1333	1	5		
/cgn2_6/prodata/2/1na/5B.COMB.seq:US-09-058-260-13		7.00	103.91	423.27	1699	1	5		
/cgn2_6/prodata/2/1na/6A.COMB.seq:US-09-058-260-13		7.00	103.69	435.64	1753	1	5		
/cgn2_6/prodata/2/1na/6B.COMB.seq:US-09-058-260-21		7.00	103.68	436.33	1756	1	5		
/cgn2_6/prodata/2/1na/6A.COMB.seq:US-09-058-260-31		7.00	103.60	440.90	1776	1	5		
/cgn2_6/prodata/2/1na/5B.COMB.seq:US-09-058-260-23		7.00	103.13	468.26	1896	1	5		
/cgn2_6/prodata/2/1na/6B.COMB.seq:US-09-058-260-17		7.00	103.02	474.85	1925	1	5		
/cgn2_6/prodata/2/1na/6A.COMB.seq:US-09-058-260-5		7.00	102.92	480.98	1952	1	5		
/cgn2_6/prodata/2/1na/6B.COMB.seq:US-09-058-260-3		7.00	101.41	583.77	2409	1	5		
/cgn2_6/prodata/2/1na/5B.COMB.seq:US-08-894-997-49		7.00	97.67	943.33	4057	1	5		
/cgn2_6/prodata/2/1na/6A.COMB.seq:US-08-894-997-49		7.00	97.61	950.39	4090	1	5		
/cgn2_6/prodata/2/1na/6B.COMB.seq:US-08-894-997-49		7.00	97.61	950.39	4090	1	5		
/cgn2_6/prodata/2/1na/5A.COMB.seq:US-08-477-451-1		7.00	95.64	1.2e+03	5368	1	5		
/cgn2_6/prodata/2/1na/5B.COMB.seq:US-08-477-451-1		7.00	95.35	1.3e+03	5599	1	5		
/cgn2_6/prodata/2/1na/6A.COMB.seq:US-08-477-451-13		7.00	94.95	1.3e+03	5825	1	5		
/cgn2_6/prodata/2/1na/6B.COMB.seq:US-08-477-451-4		7.00	94.95	1.3e+03	5825	1	5		
/cgn2_6/prodata/2/1na/5A.COMB.seq:US-08-477-451-1		7.00	94.55	1.4e+03	6263	1	5		
/cgn2_6/prodata/2/1na/5B.COMB.seq:US-08-477-451-1		7.00	94.55	1.4e+03	6263	1	5		
/cgn2_6/prodata/2/1na/6A.COMB.seq:US-08-470-202-56		7.00	91.34	2.1e+03	9793	1	5		
/cgn2_6/prodata/2/1na/6B.COMB.seq:US-08-470-202-56		7.00	91.34	2.1e+03	9793	1	5		
/cgn2_6/prodata/2/1na/5A.COMB.seq:US-08-468-059-56		7.00	91.34	2.1e+03	9793	1	5		
/cgn2_6/prodata/2/1na/5B.COMB.seq:US-08-468-059-56		7.00	90.98	2.2e+03	10299	1	5		
/cgn2_6/prodata/2/1na/6A.COMB.seq:US-08-477-451-1		7.00	90.98	2.2e+03	10299	1	5		
/cgn2_6/prodata/2/1na/6B.COMB.seq:US-08-477-451-1		7.00	90.98	2.2e+03	10299	1	5		

```
/cgn2_6/prodata/2/1na/5B.COMB.seq:US-08-477-451-25 + 7.00 86.24 4.1e+03 19932  
/cgn2_6/prodata/2/1na/6B.COMB.seq:US-09-422-869-1 - 7.00 79.76 9.3e+03 49136  
/cgn2_6/prodata/2/1na/6A.COMB.seq:US-09-103-840A-2 - 7.00 47.49 4.1e+05 44037  
/cgn2_6/prodata/2/1na/5B.COMB.seq:US-08-912-129A-66 - 6.00 118.57 64.61 44115  
/cgn2_6/prodata/2/1na/5A.COMB.seq:US-08-004-139B-10 - 6.00 118.20 67.73 20  
/cgn2_6/prodata/2/1na/5B.COMB.seq:US-08-811-492-10 - 6.00 118.20 67.73 20
```

```
seq_name: /cgn2_6/prodata/2/1na/5A.COMB.seq:US-08-411-706-3
```

seq_documentation_block:

```
Sequence 3, Application US/08411706  
Patent No. 5789193  
GENERAL INFORMATION:  
APPLICANT:  
APPLICANT:  
APPLICANT:  
APPLICANT:  
TITLE OF INVENTION: Increased production of secreted recombinant eukaryotic cell  
TITLE OF INVENTION: proteins by  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/411,706  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FI 92 4494  
FILING DATE: 06-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Thomas C.  
REGISTRATION NUMBER: 36,989  
REFERENCE/DOCKET INFORMATION: 32530  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 885 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: Saccharomyces cerevisiae  
STRAIN: X 2180-1B  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..885  
US-08-411-706-3  
alignment_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment_block:  
US-10-048-197-2 x US-08-411-706-3/rev ..
```

Align seg 1/1 to reverse of: US-08-411-706-3 from: 1 to: 885

12 AlaLeuLeuThrAlaLeuVal 19
|||||
779 GCACCTCTAACGCCCTTGTGGTG 756

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-318-978-15

seq_documentation_block:

Sequence 15, Application US/09318978A
Patent No. 6245526
GENERAL INFORMATION:
APPLICANT: Yue, Henry
APPLICANT: Kaser, Matthew, R.
APPLICANT: Baughn, Mariah, R.
TITLE OF INVENTION: LIPID METABOLISM TRANSCRIPTION FACTOR
FILE REFERENCE: PC-0004 US
CURRENT APPLICATION NUMBER: US/09/318, 978A
CURRENT FILING DATE: 1999-05-26
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PERL Program
SEQ ID NO 15
LENGTH: 250
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: unsure
LOCATION: 45
OTHER INFORMATION: a or g or c or t, unknown, or other
FEATURE:
NAME/KEY:
LOCATION:
OTHER INFORMATION: 701252210H1
PUBLICATION INFORMATION:
US-09-318-978-15

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x US-09-318-978-15 ..

Align seg 1/1 to: US-09-318-978-15 from: 1 to: 250

15 ThrAlaLeuValThrGly 21
|||||
85 ACAGCAGCTGTGGTRACCGGT 105

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-747-221B-1

seq_documentation_block:

Sequence 1, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6063610e1 Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610e1ember 12, 1996
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 401 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 92..400

FEATURE:

NAME/KEY: Xaa - Ile, Thr Lys or Arg

LOCATION: 218

FEATURE:

NAME/KEY: Xaa - Lys, Glu or Gln

LOCATION: 275, 329

FEATURE:

NAME/KEY: Xaa - Asn, Tyr or Asp

LOCATION: 332

US-08-747-221B-1

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x US-08-747-221B-1/rev ..

Align seg 1/1 to reverse of: US-08-747-221B-1 from: 1 to: 401

86 LeuSerLeuIleIleSerphe 92
|||||
156 TTATCATTAATCATTTCTTC 136

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-747-221B-3

seq_documentation_block:

Sequence 3, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6063610e1 Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0


```

1      INFORMATION FOR SEQ ID NO: 1
2      SEQUENCE CHARACTERISTICS
3          LENGTH: 401 nucleotides
4          TYPE: nucleic acid
5          STRANDEDNESS: single
6          TOPOLOGY: linear
7      MOLECULE TYPE: cDNA
8      FEATURE:
9          NAME/KEY: CDS
10         LOCATION: 92..400
11     FEATURE:
12         NAME/KEY: Xaa - Ile, Thr Lys or Arg
13         LOCATION: 218
14     FEATURE:
15         NAME/KEY: Xaa - Lys, Glu or Gln
16         LOCATION: 275, 329
17     FEATURE:
18         NAME/KEY: Xaa - Asn, Tyr or Asp
19         LOCATION: 332
20     US-09-005-051-1
21
22     alignment_scores:
23         Quality: 7.00          Length: 7
24         Ratio: 1.000          Gaps: 0
25         Percent Similarity: 100.000    Percent Identity: 100.000
26
27     alignment_block:
28     US-10-048-197-2 x US-09-005-051-1/rev ..
29
30     Align seg 1/1 to reverse of: US-09-005-051-1 from: 1 to: 401
31
32         86 LeuSerIleIleIleSerPhe 92
33         |||||||||
34         156 TTATCATTTATCATTTCTTC 136
35
36     seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-005-051-3
37
38     seq_documentation_block:
39         Sequence 3, Application US/09005051
40         Patent No. 6291222
41     GENERAL INFORMATION:
42         APPLICANT: Silver, Gary W.
43         APPLICANT: Misnewski, Nancy
44         TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
45         TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
46     NUMBER OF SEQUENCES: 66
47     CORRESPONDENCE ADDRESS:
48         ADDRESSEE: Carol Talkington Verser, Ph.D.
49         ADDRESSEE: Heska Corporation
50         STREET: 1825 Sharp Point Drive
51         City: Fort Collins
52         STATE: Colorado
53         COUNTRY: USA
54         ZIP: 80525
55     COMPUTER READABLE FORM:
56         MEDIUM TYPE: floppy disk
57         OPERATING SYSTEM: Windows 95
58         SOFTWARE: Wordperfect for Windows, Version 7.0
59     CURRENT APPLICATION DATA:
60         APPLICATION NUMBER: US/09/005,051
61         FILING DATE:
62     CLASSIFICATION:
63     PRIOR APPLICATION DATA:
64         APPLICATION NUMBER: 08/747,221
65         FILING DATE: NO. 6291222emher 12, 1996
66     ATTORNEY/AGENT INFORMATION:
67         NAME: Verser, Carol Talkington
68         REGISTRATION NUMBER: 37,459
69         REFERENCE/DOCKET NUMBER: FC-1
70     TELECOMMUNICATION INFORMATION:
71         TELEPHONE: 970/493-7272

```

TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-005-051-3

alignment_scores:

Quality:	7.00	Length:	7
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-10-048-197-2 x US-09-005-051-3 ..

Align seg 1/1 to: US-09-005-051-3 from: 1 to: 401

86 LeuSerLeuIleIleSerPhe 92
|||||
246 TTATCATTAATCATTTCTTC 266

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-367-198A-4

seq_documentation_block:

Sequence 4, Application US/08367198A
Patent No. 5627049
GENERAL INFORMATION:
APPLICANT: Menart, Sandrine
TITLE OF INVENTION: K. Iactis RP28 Ribosomal Protein Gene
TITLE OF INVENTION: Promoter and Use Thereof
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhone-Poulenc Rorer Inc.
STREET: 500 Arcola Road, 3C43
CITY: Collegeville
STATE: PA
COUNTRY: USA
ZIP: 19426-0107
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/367,198A
FILING DATE: 01-FEB-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR93/00695
FILING DATE: 06-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/08429
FILING DATE: 08-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Julie K.
REGISTRATION NUMBER: 38,619
REFERENCE/DOCKET NUMBER: ST92045-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610)454-3839
TELEFAX: (610)454-3808
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 549 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ANTI-SENSE: NO

US-08-367-198A-4

alignment_scores:

Quality:	7.00	Length:	7
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-10-048-197-2 x US-08-367-198A-4/rev ..

Align seg 1/1 to reverse of: US-08-367-198A-4 from: 1 to: 549

4 HisHisIleArgLeuThrIle 10
|||||
241 CATCATCATGACACTGACGATC 221

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-932-978-1

seq_documentation_block:

Sequence 1, Application US/08932978
Patent No. 5885804
GENERAL INFORMATION:
APPLICANT: Zalacain, Magdalena
APPLICANT: Brown, James R.
TITLE OF INVENTION: NOVEL PHO
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/932,978
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd Q
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: GM0100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 969 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-932-978-1

alignment_scores:

Quality:	7.00	Length:	7
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-10-048-197-2 x US-08-932-978-1/rev ..

Align seg 1/1 to reverse of: US-08-932-978-1 from: 1 to: 969

577 CATCACATCAGACTGACGATC 55

;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinlauf
;

```

: STREET: 1101 Connecticut Avenue
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 5.25 Inch, 360 Kb storage
: COMPUTER: IBM AT-compatible, 80286 processor
: OPERATING SYSTEM: MS-DOS version 5.0
: SOFTWARE: WordPerfect version 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/684,862
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/361,705
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/966,040
: FILING DATE: 30-DEC-1992
: APPLICATION NUMBER: PCT/EP91/01361
: FILING DATE: 19-JUL-1991
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1333 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: ORIGINAL SOURCE:
: ORGANISM: Agkistrodon rhodostoma
: FEATURE:
: LOCATION: 231 to 935
: OTHER INFORMATION: The coding region shown in (2)(1x)(B)
: OTHER INFORMATION: codes for the protein of SEQ ID NO: 4
: US-08-684-862-9

```

```

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

alignment_block:

US-10-048-197-2 x US-08-684-862-9/rev ..

Align seg 1/1 to reverse of: US-08-684-862-9 from: 1 to: 1333

```

      87 SerleuilelleSerpheleu 93
      ||||||||||||||||||
      1298 TCTCTATTATTCTCTTCTC 1278

```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-058-260-13

seq_documentation_block:

```

: Sequence 13, Application US/09058260B
: Patent No. 6218167
: GENERAL INFORMATION:
: APPLICANT: Allen, Larry
: APPLICANT: Atkins, John
: APPLICANT: Fomstein, Michael
: APPLICANT: Vonstein, Veronika
: APPLICANT: Demirjian, David
: APPLICANT: Casadaban, Malcolm
: TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
: FILE REFERENCE: 95-963-H
: CURRENT APPLICATION NUMBER: US/09/058,260B
: EARLIER FILING DATE: 1999-04-10
: EARLIER APPLICATION NUMBER: 60/001,995
: EARLIER FILING DATE: 1996-08-07
: EARLIER APPLICATION NUMBER: 60/009,704
: EARLIER FILING DATE: 1996-01-11
: EARLIER APPLICATION NUMBER: 60/019,580

```

```

: EARLIER FILING DATE: 1996-06-12
: EARLIER APPLICATION NUMBER: 08/694,078
: EARLIER FILING DATE: 1996-08-08
: EARLIER APPLICATION NUMBER: 08/781,802
: EARLIER FILING DATE: 1997-01-10
: EARLIER APPLICATION NUMBER: 08/827,810
: EARLIER FILING DATE: 1997-04-11
: NUMBER OF SEQ ID NOS: 37
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 13
: LENGTH: 1699
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
: OTHER INFORMATION: gene from bacteria E004
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (92)..(1594)
: US-09-058-260-13

```

```

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

alignment_block:

US-10-048-197-2 x US-09-058-260-13/rev ..

Align seg 1/1 to reverse of: US-09-058-260-13 from: 1 to: 1699

```

      51 GluileSerSerArphecly 57
      ||||||||||||||||||
      1413 GAATTAAGCTCCCGCTTCGCA 1393

```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-058-260-21

seq_documentation_block:

```

: Sequence 21, Application US/09058260B
: Patent No. 6218167
: GENERAL INFORMATION:
: APPLICANT: Allen, Larry
: APPLICANT: Atkins, John
: APPLICANT: Fomstein, Michael
: APPLICANT: Vonstein, Veronika
: APPLICANT: Demirjian, David
: APPLICANT: Casadaban, Malcolm
: TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
: FILE REFERENCE: 95-963-H
: CURRENT APPLICATION NUMBER: US/09/058,260B
: EARLIER FILING DATE: 1999-04-10
: EARLIER APPLICATION NUMBER: 60/001,995
: EARLIER FILING DATE: 1996-08-07
: EARLIER APPLICATION NUMBER: 60/009,704
: EARLIER FILING DATE: 1996-01-11
: EARLIER APPLICATION NUMBER: 60/019,580
: EARLIER FILING DATE: 1996-06-12
: EARLIER APPLICATION NUMBER: 08/694,078
: EARLIER FILING DATE: 1996-08-08
: EARLIER APPLICATION NUMBER: 08/781,802
: EARLIER FILING DATE: 1997-01-10
: EARLIER APPLICATION NUMBER: 08/827,810
: EARLIER FILING DATE: 1997-04-11
: NUMBER OF SEQ ID NOS: 37
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 21
: LENGTH: 1753
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
: OTHER INFORMATION: gene from bacteria E013

```

```
FEATURE:
; NAME/KEY: CDS
; LOCATION: (128)..(1630)
US-09-058-260-21
```

```
alignment_scores:
      Quality: 7.00      Length: 7
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000
```

alignment_block:

US-10-048-197-2 x US-09-058-260-21/rev ..

Align seg 1/1 to reverse of: US-09-058-260-21 from: 1 to: 1753

```
51 Glut1eserSeraArgphecly 57
|||||
1449 GAAATTAGCTCCGCTTCGGA 1429
```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-058-260-31

seq_documentation_block:

```
; Sequence 31, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
; APPLICANT: Allen, Larry
; APPLICANT: Alkems, John
; APPLICANT: Fongstein, Michael
; APPLICANT: Vonstein, Veronika
; APPLICANT: Demirjian, David
; APPLICANT: Casadaban, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; FILE REFERENCE: 95-963-H
; CURRENT APPLICATION NUMBER: US/09/058,260B
; EARLIER FILING DATE: 1999-04-10
; EARLIER APPLICATION NUMBER: 60/001,995
; EARLIER FILING DATE: 1996-08-07
; EARLIER APPLICATION NUMBER: 60/009,704
; EARLIER FILING DATE: 1996-01-11
; EARLIER APPLICATION NUMBER: 60/019,580
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: 08/694,078
; EARLIER FILING DATE: 1996-08-08
; EARLIER APPLICATION NUMBER: 08/781,802
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 08/827,810
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1756
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
; OTHER INFORMATION: gene from bacteria E027
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (130)..(1632)
US-09-058-260-31
```

```
alignment_scores:
      Quality: 7.00      Length: 7
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000
```

alignment_block:

US-10-048-197-2 x US-09-058-260-31/rev ..

Align seg 1/1 to reverse of: US-09-058-260-31 from: 1 to: 1756

```
51 Glut1eserSeraArgphecly 57
|||||
1451 GAAATTAGCTCCGCTTCGGA 1431
```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-058-260-23

seq_documentation_block:

```
; Sequence 23, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
; APPLICANT: Allen, Larry
; APPLICANT: Alkems, John
; APPLICANT: Fongstein, Michael
; APPLICANT: Vonstein, Veronika
; APPLICANT: Demirjian, David
; APPLICANT: Casadaban, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; FILE REFERENCE: 95-963-H
; CURRENT APPLICATION NUMBER: US/09/058,260B
; EARLIER FILING DATE: 1999-04-10
; EARLIER APPLICATION NUMBER: 60/001,995
; EARLIER FILING DATE: 1996-08-07
; EARLIER APPLICATION NUMBER: 60/009,704
; EARLIER FILING DATE: 1996-01-11
; EARLIER APPLICATION NUMBER: 60/019,580
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: 08/694,078
; EARLIER FILING DATE: 1996-08-08
; EARLIER APPLICATION NUMBER: 08/781,802
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 08/827,810
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 23
; LENGTH: 1776
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
; OTHER INFORMATION: gene from bacteria E015
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (128)..(1630)
US-09-058-260-23
```

```
alignment_scores:
      Quality: 7.00      Length: 7
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000
```

alignment_block:

US-10-048-197-2 x US-09-058-260-23/rev ..

Align seg 1/1 to reverse of: US-09-058-260-23 from: 1 to: 1776

```
51 Glut1eserSeraArgphecly 57
|||||
1449 GAAATTAGCTCCGCTTCGGA 1429
```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-058-260-3

seq_documentation_block:

```
; Sequence 3, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
; APPLICANT: Allen, Larry
; APPLICANT: Alkems, John
; APPLICANT: Fongstein, Michael
; APPLICANT: Vonstein, Veronika
; APPLICANT: Demirjian, David
; APPLICANT: Casadaban, Malcolm
```

```
;; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;; FILE REFERENCE: 95-963-H
;; CURRENT APPLICATION NUMBER: US/09/058,260B
;; EARLIER FILING DATE: 1999-04-10
;; EARLIER APPLICATION NUMBER: 60/001,995
;; EARLIER FILING DATE: 1996-08-07
;; EARLIER APPLICATION NUMBER: 60/009,704
;; EARLIER FILING DATE: 1996-01-11
;; EARLIER APPLICATION NUMBER: 60/019,580
;; EARLIER FILING DATE: 1996-06-12
;; EARLIER APPLICATION NUMBER: 08/694,078
;; EARLIER FILING DATE: 1996-08-08
;; EARLIER APPLICATION NUMBER: 08/781,802
;; EARLIER FILING DATE: 1997-01-10
;; EARLIER APPLICATION NUMBER: 08/827,810
;; EARLIER FILING DATE: 1997-04-11
;; NUMBER OF SEQ ID NOS: 37
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 3
;; LENGTH: 1896
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;; OTHER INFORMATION: gene from bacteria E009
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (211)..(1713)
;; US-09-058-260-3
```

```
alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

alignment block:

US-10-048-197-2 x US-09-058-260-3/rev ..

Align seg 1/1 to reverse of: US-09-058-260-3 from: 1 to: 1896

```
51 GlutlSerSerArgphecly 57
|||||
1532 GAATTAGCTCCCGCTTCGGA 1512
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-17

seq_documentation_block:

```
;; Sequence 17, Application US/09058260B
;; Patent No. 6218167
;; GENERAL INFORMATION:
;; APPLICANT: Allen, Larry
;; APPLICANT: Aikens, John
;; APPLICANT: Fonstein, Michael
;; APPLICANT: Vonstein, Veronika
;; APPLICANT: Demirjian, David
;; APPLICANT: Casadaban, Malcolm
;; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;; FILE REFERENCE: 95-963-H
;; CURRENT APPLICATION NUMBER: US/09/058,260B
;; CURRENT FILING DATE: 1999-04-10
;; EARLIER APPLICATION NUMBER: 60/001,995
;; EARLIER FILING DATE: 1996-08-07
;; EARLIER APPLICATION NUMBER: 60/009,704
;; EARLIER FILING DATE: 1996-01-11
;; EARLIER APPLICATION NUMBER: 60/019,580
;; EARLIER FILING DATE: 1996-06-12
;; EARLIER APPLICATION NUMBER: 08/694,078
;; EARLIER FILING DATE: 1996-08-08
;; EARLIER APPLICATION NUMBER: 08/781,802
;; EARLIER FILING DATE: 1997-01-10
;; EARLIER APPLICATION NUMBER: 08/827,810
;; EARLIER FILING DATE: 1997-04-11
```

```
;; NUMBER OF SEQ ID NOS: 37
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 17
;; LENGTH: 1925
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;; OTHER INFORMATION: gene from bacteria E008
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (127)..(1581)
;; US-09-058-260-17
```

```
alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

alignment block:

US-10-048-197-2 x US-09-058-260-17/rev ..

Align seg 1/1 to reverse of: US-09-058-260-17 from: 1 to: 1925

```
51 GlutlSerSerArgphecly 57
|||||
1400 GAATTAGCTCCCGCTTCGGA 1380
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-5

seq_documentation_block:

```
;; Sequence 5, Application US/09058260B
;; Patent No. 6218167
;; GENERAL INFORMATION:
;; APPLICANT: Allen, Larry
;; APPLICANT: Aikens, John
;; APPLICANT: Fonstein, Michael
;; APPLICANT: Vonstein, Veronika
;; APPLICANT: Demirjian, David
;; APPLICANT: Casadaban, Malcolm
;; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;; FILE REFERENCE: 95-963-H
;; CURRENT APPLICATION NUMBER: US/09/058,260B
;; CURRENT FILING DATE: 1999-04-10
;; EARLIER APPLICATION NUMBER: 60/001,995
;; EARLIER FILING DATE: 1996-08-07
;; EARLIER APPLICATION NUMBER: 60/009,704
;; EARLIER FILING DATE: 1996-01-11
;; EARLIER APPLICATION NUMBER: 60/019,580
;; EARLIER FILING DATE: 1996-06-12
;; EARLIER APPLICATION NUMBER: 08/694,078
;; EARLIER FILING DATE: 1996-08-08
;; EARLIER APPLICATION NUMBER: 08/781,802
;; EARLIER FILING DATE: 1997-01-10
;; EARLIER APPLICATION NUMBER: 08/827,810
;; EARLIER FILING DATE: 1997-04-11
;; NUMBER OF SEQ ID NOS: 37
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 5
;; LENGTH: 1952
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;; OTHER INFORMATION: gene from bacteria E011
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (197)..(1699)
;; US-09-058-260-5
```

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x US-09-058-260-5/rev ..

Align seg 1/1 to reverse of: US-09-058-260-5 from: 1 to: 1952

51 GiuleSerSerArgphecly 57

1518 GAATTAGCTCCGCTCGA 1498

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-263-023-3

seq_documentation_block:

Sequence 3, Application US/09263023
Patent No. 6037159
GENERAL INFORMATION:
APPLICANT: Uchimura, Kenji
APPLICANT: Muramatsu, Hideki
APPLICANT: Kadomatsu, Kenji
APPLICANT: Kanagaki, Reiji
APPLICANT: Habuchi, Osami
APPLICANT: Muramatsu, Takashi
TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
TITLE OF INVENTION: DNA ENCODING THE SAME
FILE REFERENCE: TOYAMA1.001AUS
CURRENT APPLICATION NUMBER: US/09/263.023
CURRENT FILING DATE: 1999-03-05
EARLIER APPLICATION NUMBER: JP 10-54007
EARLIER FILING DATE: 1998-03-05
EARLIER APPLICATION NUMBER: JP 10-177844
EARLIER FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 2409
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (390)...(1841)
US-09-263-023-3

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x US-09-263-023-3 ..

Align seg 1/1 to: US-09-263-023-3 from: 1 to: 2409

11 SerAlaLeuLeuThAlaLeu 17

614 TCCGACTCTTACCGCCCTC 634

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-894-997-49

seq_documentation_block:

Sequence 49, Application US/08894997A
Patent No. 6270990
GENERAL INFORMATION:
APPLICANT: Anderson, David J
APPLICANT: Schoenheit, Christopher J
TITLE OF INVENTION: NEURON-RESTRICTIVE SILENCER FACTOR
FILE REFERENCE: 17810-502 NRSF
CURRENT APPLICATION NUMBER: US/08/894.997A
CURRENT FILING DATE: 1998-01-06
EARLIER APPLICATION NUMBER: PCT/US96/02817

EARLIER FILING DATE: 1996-03-01
EARLIER APPLICATION NUMBER: 08/298,590
EARLIER FILING DATE: 1995-03-03
NUMBER OF SEQ ID NOS: 55
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 49
LENGTH: 4057

TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: gene
LOCATION: (1)...(4057)
OTHER INFORMATION: Human NSRF
US-08-894-997-49

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-048-197-2 x US-08-894-997-49/rev ..

Align seg 1/1 to reverse of: US-08-894-997-49 from: 1 to: 4057

44 LysGlyLysThrAsnLysGln 50

4011 AAGGAAAAACAAACAAACAA 3991

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-781-802-5

seq_documentation_block:

Sequence 5, Application US/08781802
Patent No. 5969121
GENERAL INFORMATION:
APPLICANT: ALLEN, Larry
APPLICANT: ATKENS, John
APPLICANT: FONSTEIN, Michael
APPLICANT: VONSTEIN, Veronika
APPLICANT: DEMIRJIAN, David
APPLICANT: CASADABAN, Malcolm
TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 S. Wacker Drive 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781.802
FILING DATE: 10-JAN-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/694,078
FILING DATE: 07-AUG-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/019,580
FILING DATE: 12-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/009,704
FILING DATE: 11-JAN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/001,995
FILING DATE: 01-AUG-1995

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 95,963-E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4090 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 197..1699
; OTHER INFORMATION: /note="E011 sequence of longest
; OTHER INFORMATION: open reading frame; other possible start codons ATG/met4;
; OTHER INFORMATION: TTG/Leu7; GTG/Val8; GTG/Val15; GTG/Val36; ATG/met62"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 197..1699
; US-08-781-802-5

alignment_scores:
    Quality: 7.00      Length: 7
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-781-802-5/rev ..
Align seg 1/1 to reverse of: US-08-781-802-5 from: 1 to: 4090
51 GlutSerSerArgpHeGly 57
|||||
1518 GAAATTAGCTCCCGCTCGGA 1498

seq_name: /cgn2_6/prodata/2/ina/5B_COMB.seq:US-08-694-078-5
seq_documentation_block:
; Sequence 5, Application US/08694078
; Patent No. 6218163
; GENERAL INFORMATION:
; APPLICANT: ALLEN, Larry
; APPLICANT: AIKENS, John
; APPLICANT: FONSESTEIN, Michael
; APPLICANT: VONSTEIN, Veronika
; APPLICANT: DEMIRJIAN, David
; APPLICANT: CASADABAN, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff, Ltd.
; STREET: 300 S. Wacker Drive 7th Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,078
; FILING DATE: 07-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/019,580
; FILING DATE: 12-JUN-1996
;
;
```

```

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/009,704
; FILING DATE: 10-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/001,995
; FILING DATE: 07-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 95,963-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4090 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 197..1699
; OTHER INFORMATION: /note="E011 sequence of longest
; OTHER INFORMATION: open reading frame; other possible start codons ATG/met4;
; OTHER INFORMATION: TTG/Leu7; GTG/Val8; GTG/Val15; GTG/Val36; ATG/met62"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 197..1699
; US-08-694-078-5

alignment_scores:
    Quality: 7.00      Length: 7
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-694-078-5/rev ..
Align seg 1/1 to reverse of: US-08-694-078-5 from: 1 to: 4090
51 GlutSerSerArgpHeGly 57
|||||
1518 GAAATTAGCTCCCGCTCGGA 1498

seq_name: /cgn2_6/prodata/2/ina/5B_COMB.seq:US-08-485-355B-45
seq_documentation_block:
; Sequence 45, Application US/08485355B
; Patent No. 6177075
; GENERAL INFORMATION:
; APPLICANT: CHRISTIAN, P. D., Gordon, K. H.J., Hanzlik, T. N.
; TITLE OF INVENTION: Insect Viruses and Their Uses in
; Protecting Plants
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,355B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
;
;
```


APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-2/RTT/DSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 272729
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 5368 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 4944..5162
SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-08-485-355B-45

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-485-355B-45 ..
Align seg 1/1 to: US-08-485-355B-45 from: 1 to: 5368

30 Mettlysglucinsnglncin 36
|||||
5190 ATGAGGACAAATCAGCAG 5210

seq_name: /cgn2_6/ptodata/2/ina/Backfiles1.seq:5223424-5

seq_documentation_block:

Patent No. 5223424
APPLICANT: COCHRAN, MARK;CHIANG, CHRISTINA H.;MACDONALD,
RICHARD D.
TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND
HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO
ACID SEQUENCE
NUMBER OF SEQUENCES: 16
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/225,032
FILING DATE: 27-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 78,519
FILING DATE: 27-JUL-1987
APPLICATION NUMBER: 933,107
FILING DATE: 20-NOV-1986
APPLICATION NUMBER: 902,887
FILING DATE: 02-SEP-1986
APPLICATION NUMBER: 887,140
FILING DATE: 17-JUL-1986
APPLICATION NUMBER: 823,102
FILING DATE: 27-JAN-1986
APPLICATION NUMBER: 773,430
FILING DATE: 06-SEP-1985
SEQ ID NO: 5:
LENGTH: 5379
5223424-5

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x 5223424-5/rev ..

Align seg 1/1 to reverse of: 5223424-5 from: 1 to: 5379

8 LeuthThilesSerlaleuleu 14
|||||
733 TTAACATAAGTCGCTTCTTA 713

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-477-451-9

seq_documentation_block:

Sequence 9, Application US/08477451
Patent No. 5928865
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335,002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 5599 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-451-9

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-477-451-9 ..

Align seg 1/1 to: US-08-477-451-9 from: 1 to: 5599

85 lileuSerleuileSer 91
|||||
3333 ATTTCAGCTGTGATCATTTTCG 3353

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-477-451-13

seq_documentation_block:

```
; Sequence 13, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covaccl, Antonello
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5599 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-477-451-13
```

```
alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

alignment_block:
US-10-048-197-2 x US-08-477-451-13/rev ..

Align seg 1/1 to reverse of: US-08-477-451-13 from: 1 to: 5599

```
85 lleuSerleuIleIleSer 91
|||||
2267 ATTGAGCTTGATCATTTGCG 2247

seq_name: /cgn2_6/ptcodata/2/ina/6a_COMB.seq:US-08-470-260-4
```

```
seq_documentation_block:
; Sequence 4, Application US/08470260
; Patent No. 6077706
; GENERAL INFORMATION:
; APPLICANT: Covaccl, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
; TITLE OF INVENTION: for Vaccines and Diagnostics
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
```

5

```
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,260
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,848
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0316.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5925 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-470-260-4
```

```
alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

alignment_block:
US-10-048-197-2 x US-08-470-260-4 ..

Align seg 1/1 to: US-08-470-260-4 from: 1 to: 5925

```
90 lleSerPheIleuTrpVallys 96
|||||
4801 ATCTCATTTTGTGGGTAAAA 4821
```

seq_name: /cgn2_6/ptcodata/2/ina/6a_COMB.seq:US-08-471-491-4
seq_documentation_block:

```
; Sequence 4, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
; APPLICANT: Covaccl, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; TITLE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0044
; CURRENT APPLICATION NUMBER: US/08/471,491B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 5925
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; US-08-471-491-4
```

```
alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

alignment_block:
US-10-048-197-2 x US-08-471-491-4 ..
Align seg 1/1 to: US-08-471-491-4 from: 1 to: 5925
90 IleserphelLeuTrpVallys 96
|||||
4801 ATCTCATTTTGTGGGTAAAA 4821

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-466-662-4
seq_documentation_block:
; Sequence 4, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; FILE REFERENCE: CHIR0057
; CURRENT APPLICATION NUMBER: US/08/466, 662B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 5925
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-08-466-662-4

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-466-662-4 ..
Align seg 1/1 to: US-08-466-662-4 from: 1 to: 5925
90 IleserphelLeuTrpVallys 96
|||||
4801 ATCTCATTTTGTGGGTAAAA 4821

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-781-802-3
seq_documentation_block:
; Sequence 3, Application US/08781802
; Patent No. 5969121
; GENERAL INFORMATION:
; APPLICANT: ALLEN, Larry
; APPLICANT: AIKENS, John
; APPLICANT: FONSTEIN, Michael
; APPLICANT: VONSTEIN, Veronika
; APPLICANT: DEMIRJIAN, David
; APPLICANT: CASADABAN, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 S. Wacker Drive 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,802
FILING DATE: 10-JAN-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/694,078
FILING DATE: 07-AUG-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/019,580
FILING DATE: 12-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/009,704
FILING DATE: 11-JAN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/001,995
FILING DATE: 01-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Chao, Mark
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 95,963-E
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6263 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 2314..3816
OTHER INFORMATION: /note= "E009 sequence with longest
OTHER INFORMATION: open reading frame; possible other start codons are ATG/me
OTHER INFORMATION: TTG/Leu7; GTG/Val8; GTG/Val15; GTG/Val36"
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 2314..3816
US-08-781-802-3

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-781-802-3/rev ..
Align seg 1/1 to reverse of: US-08-781-802-3 from: 1 to: 6263
51 GIUileserSerArpGhegly 57
|||||
3635 GAATTAGCTCCGCTTCGGA 3615

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-694-078-3
seq_documentation_block:
; Sequence 3, Application US/08694078
; Patent No. 6218163
; GENERAL INFORMATION:
; APPLICANT: ALLEN, Larry
; APPLICANT: AIKENS, John
; APPLICANT: FONSTEIN, Michael
; APPLICANT: VONSTEIN, Veronika
; APPLICANT: DEMIRJIAN, David
; APPLICANT: CASADABAN, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; NUMBER OF SEQUENCES: 8

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: McDONNELL BOEHLEN HULBERT & BERGHOFF, LTD.
STREET: 300 S. WACKER DRIVE 7TH FLOOR
CITY: CHICAGO
STATE: ILLINOIS
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/694,078
FILING DATE: 07-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/019,580
FILING DATE: 12-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/009,704
FILING DATE: 10-JAN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/001,995
FILING DATE: 07-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Chao, Mark
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 95,963-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6263 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 2314..3816
OTHER INFORMATION: /note="E009 sequence with longest
open reading frame; possible other start codons are ATG/mel4,
OTHER INFORMATION: TTG/leu7; GTG/val8; GTG/val15; GTG/val36"
FEATURE:
NAME/KEY: mat.peptide
LOCATION: 2314..3816
US-08-694-078-3

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-694-078-3/rev ..

Align seg 1/1 to reverse of: US-08-694-078-3 from: 1 to: 6263

51 GIUIISeSerArphecly 57
|||||
3635 GAAATTAGCTCCGCTTCGGA 3615

seq_name: /cgn2_6/plodata/2/ina/5A_COMB.seq:US-08-619-554-1

seq_documentation_block:
; Sequence 1, Application US/08619554
; Patent No. 5821353
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS, Cameron M.
; APPLICANT: CHREBET, Gary L.
3

```

```

1  APPLICANT: CLEMAS, Joseph
2  APPLICANT: EL-SHERBINI, Mohammed
3  APPLICANT: FOOR, Forrest
4  APPLICANT: KAHN, Jennifer
5  APPLICANT: KELLY, Rosemarie, - PARENT, S.A.
6  APPLICANT: MORRINAN, Jean, - RAHADAN, N.M.
7  APPLICANT: MORIN, Nancy, - REGISTER, E.A.
8  APPLICANT: ONISHI, Janet, - SHEI, Gan-Tu
9  TITLE OF INVENTION: DNA ENCODING 1,3 BETA-D GLUCAN
10 TITLE OF INVENTION: SYNTHASE SUBUNITS
11 NUMBER OF SEQUENCES: 8
12 CORRESPONDENCE ADDRESS:
13 ADDRESSEE: JOSEPH A. COPPOLA - MERCK & CO., INC.
14 STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
15 CITY: RAHWAY
16 STATE: NJ
17 COUNTRY: USA
18 ZIP: 07065
19 COMPUTER READABLE FORM:
20 MEDIUM TYPE: Diskette
21 COMPUTER: IBM Compatible
22 OPERATING SYSTEM: DOS
23 SOFTWARE: FASTSEQ for Windows Version 2.0
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/08/619,554
26 FILING DATE: 01-AUG-1996
27 CLASSIFICATION: 536
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER:
30 FILING DATE:
31 ATTORNEY/AGENT INFORMATION:
32 NAME: COPPOLA, JOSEPH A
33 REGISTRATION NUMBER: 38,413
34 REFERENCE/DOCKET NUMBER: 19104PI
35 TELECOMMUNICATION INFORMATION:
36 TELEPHONE: 732-594-6734
37 TELEFAX: 732-594-4720
38 TELEX:
39 INFORMATION FOR SEQ ID NO: 1:
40 SEQUENCE CHARACTERISTICS:
41 LENGTH: 7655 base pairs
42 TYPE: nucleic acid
43 STRANDEDNESS: single
44 TOPOLOGY: linear
45 MOLECULE TYPE: cDNA
46 US-08-619-554-1
47
48 alignment_scores:
49 Quality: 7.00 Length: 7
50 Ratio: 1.000 Gaps: 0
51 Percent Similarity: 100.000 Percent Identity: 100.000
52
53 alignment_block:
54 US-10-048-197-2 x US-08-619-554-1 ..
55
56 Align seg 1/1 to: US-08-619-554-1 from: 1 to: 7655
57
58 12 AlaLeuEuThAlaLeuEu 18
59 |||||||||||||||||||
60 496 GCCCTCCTTACTGCTTCCTT 516
61
62 seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-470-202-56
63
64 seq_documentation_block:
65 Sequence 56, Application US/08470202
66 Patent NO. 5759808
67 GENERAL INFORMATION:
68 APPLICANT: Guertler, Iutz G.
69 APPLICANT: Eberler, Josef
70 APPLICANT: Brunn, Albrecht v.
71 APPLICANT: Knapp, Stefan
72 APPLICANT: Hauser, Hans-Peter

```

```

? TITLE OF INVENTION: Retrovirus from the HIV Group and Its
? TITLE OF INVENTION: Use
? NUMBER OF SEQUENCES: 63
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
? ADDRESSEE: Dunner
? STREET: 1300 I Street, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: USA
? ZIP: 20005
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/470,202
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/132,653
? FILING DATE: 05-OCT-1993
? APPLICATION NUMBER: DE P 42 33 646.5
? FILING DATE: 06-OCT-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: DE P 42 35 718.7
? FILING DATE: 22-OCT-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: DE P 42 44 541.8
? FILING DATE: 30-DEC-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: DE P 43 18 186.4
? FILING DATE: 01-JUN-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Michael J. Blake
? REGISTRATION NUMBER: 37,096
? REFERENCE/DOCKET NUMBER: 05495-0001-00000
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-408-4000
? TELEFAX: 202-408-4400
? INFORMATION FOR SEQ ID NO: 56:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 9793 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-08-470-202-56

alignment_scores:
? Quality: 7.00 Length: 7
? Ratio: 1.000 Gaps: 0
? Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-470-202-56/rev ..
Align seg 1/1 to reverse of: US-08-470-202-56 from: 1 to: 9793

11 SerAlaLeuThrAlaLeu 17
|||||
6097 TCAGCGTACTTACTGCTCTG 6077

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-471-770-56
seq_documentation_block:
? Sequence 56, Application US/08471770
? Patent No. 5770427
? GENERAL INFORMATION:
? APPLICANT: Guerlier, Lutz G.
? APPLICANT: Eberle, Josef
```

```

? APPLICANT: Brunn, Albrecht V.
? APPLICANT: Knapp, Stefan
? APPLICANT: Hauser, Hans-Peter
? TITLE OF INVENTION: Retrovirus from the HIV Group and Its
? TITLE OF INVENTION: Use
? NUMBER OF SEQUENCES: 63
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
? ADDRESSEE: Dunner
? STREET: 1300 I Street, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: USA
? ZIP: 20005
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/471,770
? FILING DATE: 06-JUN-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/132,653
? FILING DATE: 05-OCT-1993
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: DE P 42 33 646.5
? FILING DATE: 06-OCT-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: DE P 42 35 718.7
? FILING DATE: 22-OCT-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: DE P 42 44 541.8
? FILING DATE: 30-DEC-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: DE P 43 18 186.4
? FILING DATE: 01-JUN-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Carol P. Binaudi
? REGISTRATION NUMBER: 32,220
? REFERENCE/DOCKET NUMBER: 05495-0001-03000
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-408-4000
? TELEFAX: 202-408-4400
? INFORMATION FOR SEQ ID NO: 56:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 9793 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-08-471-770-56

alignment_scores:
? Quality: 7.00 Length: 7
? Ratio: 1.000 Gaps: 0
? Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-471-770-56/rev ..
Align seg 1/1 to reverse of: US-08-471-770-56 from: 1 to: 9793

11 SerAlaLeuThrAlaLeu 17
|||||
6097 TCAGCGTACTTACTGCTCTG 6077

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-468-059-56
seq_documentation_block:
```

```

Sequence 56, Application US/08468059
Patent No. 5840480
GENERAL INFORMATION:
APPLICANT: Guerlier, Lutz G.
APPLICANT: Eberle, Josef
APPLICANT: Brunn, Albrecht V.
APPLICANT: Knapp, Stefan
APPLICANT: Hauser, Hans-Peter
TITLE OF INVENTION: Retrovirus from the HIV Group and Its
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunneer
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,059
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/132,653
FILING DATE: 05-OCT-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 42 33 646.5
FILING DATE: 06-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 42 35 718.7
FILING DATE: 22-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 42 44 541.8
FILING DATE: 30-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 43 18 186.4
FILING DATE: 01-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Carol P. Einaudi
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 05495-0001-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 9793 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-468-059-56

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-468-059-56/rev ..
Align seg 1/1 to reverse of: US-08-468-059-56 from: 1 to: 9793
11 SerialNumberThraLeu 17
|||||

```

```

6097 TCAGGTTACTTACTGCTCTG 6077
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-09-109-916-56
seq_documentation_block:
Sequence 56, Application US/09109916
Patent No. 627561
GENERAL INFORMATION:
APPLICANT: Guerlier, Lutz G.
APPLICANT: Eberle, Josef
APPLICANT: Brunn, Albrecht V.
APPLICANT: Knapp, Stefan
APPLICANT: Hauser, Hans-Peter
TITLE OF INVENTION: RETROVIRUS FROM HIV GROUP AND ITS USE
FILE REFERENCE: 05495.0001-04
CURRENT APPLICATION NUMBER: US/09/109,916
CURRENT FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: DE P 42 33 646.5
EARLIER FILING DATE: 1992-10-06
EARLIER APPLICATION NUMBER: DE P 42 35 718.7
EARLIER FILING DATE: 1992-10-22
EARLIER APPLICATION NUMBER: DE P 42 44 541.8
EARLIER FILING DATE: 1992-12-30
EARLIER APPLICATION NUMBER: DE P 43 18 186.4
EARLIER FILING DATE: 1993-06-01
NUMBER OF SEQ ID NOS: 67
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 56
LENGTH: 9793
TYPE: DNA
ORGANISM: Human immunodeficiency virus
US-09-109-916-56

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-09-109-916-56/rev ..
Align seg 1/1 to reverse of: US-09-109-916-56 from: 1 to: 9793
11 SerialNumberThraLeu 17
|||||
6097 TCAGGTTACTTACTGCTCTG 6077
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-477-451-1
seq_documentation_block:
Sequence 1, Application US/08477451
Patent No. 5928665
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435

```

ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10299 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-451-1

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-477-451-1 ..

Align seg 1/1 to: US-08-477-451-1 from: 1 to: 10299

90 IleserPheleutrypallys 96
|||||
9175 ATCTCATTTTGTGGGTAA 9195

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-477-451-5

seq_documentation_block:
Sequence 5, Application US/08477451
Patent No. 5928865
GENERAL INFORMATION:
APPLICANT: Covaccl, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10299 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-451-5

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-477-451-5/rev ..

Align seg 1/1 to reverse of: US-08-477-451-5 from: 1 to: 10299

90 IleserPheleutrypallys 96
|||||
1125 ATCTCATTTTGTGGGTAA 1105

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-477-451-25

seq_documentation_block:
Sequence 25, Application US/08477451
Patent No. 5928865
GENERAL INFORMATION:
APPLICANT: Covaccl, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 19932 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-451-25

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-477-451-25 ..

Align seg 1/1 to: US-08-477-451-25 from: 1 to: 19932

85 IleserPheleutrypallys 91
|||||
6890 ATTTGAGCTTGATCATTCG 6910

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-422-869-1

seq_documentation_block:

```

; Sequence 1, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANTS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 1
; LENGTH: 49136
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-1

```

```

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

alignment_block:

US-10-048-197-2 x US-09-422-869-1/rev ..

Align seq 1/1 to reverse of: US-09-422-869-1 from: 1 to: 49136

```

      8 leuThrIleSerAlaLeuIeu 14
      ||||||||||||||||
46568 CTACCATATCAGCCCTCTG 46548

```